An Empirical Study of Re-sampling Techniques as a Method for Improving Error Estimates in Split-plot Designs

THESIS

Benjamin M. Lee, Captain, USAF

AFIT-OR-MS-ENS-10-06

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-OR-MS-ENS-10-06

An Empirical Study of Re-sampling Techniques as a Method for Improving Error Estimates in Split-plot Designs

THESIS

Presented to the Faculty

Department of Operations Research

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Operations Research

Benjamin M. Lee, B.S.

Captain, USAF

March 2010

AFIT-OR-MS-ENS-10-06

An Empirical Study of Re-sampling Techniques as a Method for Improving Error Estimates in Split-plot Designs

Benjamin M. Lee, B.S.

Captain, USAF

Approved:

/signed/	22 March 2010
Dr. R. R. Hill, PhD (Chairman)	date
/signed/	22 March 2010
Maj S. R. Capehart, PhD (Member)	date

AFIT-OR-MS-ENS-10-06

Abstract

For any acquisition program, whether Department of Defense (DOD) or industry related, the primary driving factor behind the success of a program is whether or not the program remains within budget, stays on schedule and meets the defined performance requirements. If any of these three criteria are not met, the program manager may need to make challenging decisions. Typically, if the program is expected to not stay within budget or is expected to be delayed for one reason or another, the program manager will tend to limit areas of testing in order to meet these criteria. The result tends to be a reduction in the test budget and/or a shortening in the test timeline, both of which are already lean. The T&E community needs new test methodologies to test systems and gain insight on whether a system meets performance standards, within the budget and timeline constraints. In particular, both fundamental and advanced aspects of experimental design need to be adapted.

The use of experiential design within DOD has continued to grow because of the needed adaptation. Many different types of experiments have been used. An experimental design that is often needed is one that involves a restricted randomization design such as a split-plot design. Split-plot designs arise when specific factors are difficult (or impossible) to vary, a frequent occurrence within the T&E community. However, split-plot designs have limitations on the estimation of the whole-plot (hard-to-change) and subplot (easier-to-change) errors without the conduct of a sufficient number of replications for the design. Within the timeline constraints for particular programs, sufficient replications are difficult, even impossible to complete. The inability to conduct the sufficient replications often lead to models that lack precision in error estimation and thus imprecision in corresponding conclusions.

This work develops and examines a methodology for analyzing test results conducted by split-plot designs using re-sampling techniques to provide better estimates

of the error terms. The premise is to determine a set of rules using bootstrapping, a particular re-sampling technique, that can be applied to the analysis of a split-plot design, in order to create a representative regression model that can be used by the T&E community to gain required system insight.

Preface

This work is dedicated to all who gave and continue to give in order for me to achieve some semblance of success.

Benjamin M. Lee

Acknowledgements

I would like to personally thank all of those who contributed greatly to the completion of this research. First and foremost, God, without whom I would not be able to accomplish anything. Second, those in the academic environment, in particular, Dr. Hill, without his guidance and instruction this effort would not have been completed. Third, but not least, my wife and kids who gave up the most in this academic endeavor. I pray that it was well worth the hardship it may have cost them.

Benjamin M. Lee

Table of Contents

Page

List of Figures

List of Tables

Table Page

Table Page

Table		Page

Table		Page

Table Page

Table		Page

Table Page

Table Page

An Empirical Study of Re-sampling Techniques as a Method for Improving Error Estimates in Split-plot Designs

I. Introduction

1.1 Background

Cost, schedule and performance typically drive the decisions that program managers make in a system acquisition lifecycle, whether Department of Defense (DOD) or industry related. In fact, the program manager's success is generally defined by how well the program stays under cost, stays within scheduled time constraints, and meets pre-determined performance objectives. Within the DOD, the program manager's success is handicapped by limited budgets, highly technical requirements and immediate warfighter operational requirements. Thus, the program manager is nearly always in a highly stressed environment, mitigating risk, trying to stay under cost, stay within schedule and meet sometimes dynamic performance objectives. In many cases a program manager uses a reduction in Test and Evaluation (T&E) as a potential solution. However, T&E is a crucial part of the Defense Acquisition and Management System. In fact, T&E needs to provide accurate and relevant assessments of system performance and provide early identification of any deficiencies which allow for corrective actions to take place. The test community needs the ability to make statistical assertions based on test results to best meet acquisition program needs.

Clearly, the ability for T&E to provide accurate and relevant assessments, provide early identification of problems, and make valid statistical assertions is greatly impacted by any forced reduction in the test effort. Adverse outcomes of reduced test efforts include: the system may not be fully tested, not enough test conditions are used to generate statistical confidence and power, the tester is unable to identify and understand the system-under-test (SUT) in order to fix problems. These outcomes,

and others, highlight the need to use test resources efficiently and effectively. In fact, the use of industry best practices and state-of-the-art statistical methodologies may improve the ability of the T&E enterprise [7]. These best practices and methodologies are entering the DOD test community with the advent of an emphasis on experimental design practice and are helping address the impacts of limited testing. There are, however, still limitations and additional analytical advancements needed.

This research is an empirical study examining statistical methodologies with potential applicability in improving the analytical results of certain experimental designs, split-plot designs. These results could be applied by the T&E community to better obtain several objectives pertinent to T&E: mitigate risk for fielding the system, improve system performance by fully understanding the SUT, ensure the system meets operational requirements and limit total cost of test.

1.2 Problem Statement

Completely randomized designs (CRDs), such as factorial and fractional factorial designs, have been the popular method to plan and conduct tests within DOD T&E. These designs focus on various combinations of factor settings and complete randomization of the schedule of experimental runs, which is ideal. Unfortunately, complete randomization, sometimes referred to as "full randomization," is sometimes neither feasible nor effective. For instance, there may be a hard-to-change or costly factor(s) whose randomization would hurt the test conduct efficiency. Therefore, a restricted randomization approach is utilized, such as a split-plot design. Unlike standard statistical models, split-plot designs involve two types of experimental error, whole-plot and subplot error. The whole-plot is associated with the hard-to-change factors while the subplot error is associated with the fully randomized, or easy-to-change, factors. To estimate the whole-plot error, design replications are needed. However, resources often do not allow for sufficient replication. In this case, the experimenters may not be able to determine if the non-randomized whole-plot factor had a treatment effect, or even get an accurate representation of the whole-plot error.

This leads to the question: Is there a method(s) of analysis that could be applied to the non-randomized factor to determine the treatment effect and provide a more reasonable estimate for the whole-plot error?

Because of the inability to perform replicates or many multiple replicates of a split-plot design, the number of degrees of freedom for the whole-plot error is relatively small. More replicates means more degrees of freedom for the whole-plot error, thereby increasing the precision of the test. Thus, is there a method(s) that can compensate for the small degrees of freedom associated with whole-plot error; thereby increasing the precision of the test without conducting more test points?

1.3 Research Objectives/Questions/Hypotheses

The research objective is to develop, examine and test methodologies for analyzing test results from split-plot designs. In particular, this work determines the applicability of bootstrapping in supporting the analysis of split-plot designs. A determination on when bootstrapping is effective is made. The research inspects an array of split-plot design models and approaches.

1.4 Research Focus

1.4.1 Methodology. There are cases in which split-plot designs are more suitable than other experimental designs, due to restrictions on randomization. Kowalski and Potcner [24] state, in regards to CRDs,

> In practice, however, the limitations and challenges of experimenting in the real world result in these simple experiments being the exception rather than the norm. Typically, an experiment will contain some form of a restriction on the randomization. [24]

There are cases in which a CRD is unrealistic and the split-plot design will result in considerable experimental efficiency. A split-plot design does have certain limitations it presents in the analysis. For example, " whole-plot treatments in a split-plot design are confounded with the whole-plots and the subplot treatments are not

confounded, it is best to assign the factor we are most interested in to the subplots, if possible" [33]; The effect of the whole-plot factor, which will have the least number of experimental replicates, is estimated less precisely than the subplot factors, which will have more experimental replicates [24]. Therefore, this research examines the merit of using re-sampling techniques, in particular bootstrapping, as a method for increasing the precision of the whole-plot error estimates in split-plot designs. This is done via an empirical study with *a priori* split-plot design models, beginning with the most simple case and progressing to more complex, where the whole-plot and subplot errors are " known ".

1.4.2 Assumptions/Limitations. The assumptions for this research are the following:

1. The regression model used to generate the initial samples is a good representation of the true model for the system-under-test.
2. Bootstrapping can be applied to a small sample size with reasonable accuracy (bootstrapping does not necessarily work well with small sample sizes).
3. Guidelines for using bootstrapping techniques can be generated as a result of this empirical study.

1.4.3 Implications. This research has implications on the T&E testing community. CRDs are the exception, not the norm for testing [24]. This implies that split-plot designs, and other non-completely randomized, designs are used more frequently. If a "true" performance model could be represented with even fewer test points, and/or from a typical split-plot design with the use of re-sampling techniques, it could greatly benefit the T&E community. Application of DOE already creates a potential reduction in test time, provides more insight on system performance and is a potential cost reduction, due to conducting fewer test points. If there was a way to increase the benefit and gain more insight with fewer test points for a particular test, this may increase the number or types of tests performed in a test program.

1.5 Preview

This research is an empirical study of re-sampling techniques and the impacts that these techniques have on the analysis of split-plot designs. Chapter II, Literary Review, summarizes the literature background for the research. Included in this chapter are the following topics: background on the role of T&E, particularly within the DOD; a historical account of experimental design and how it has changed the face of T&E, focused mostly on split-plot designs; and a description and definition of particular re-sampling techniques, in particular, the bootstrap method. Chapter III, Methodology, provides the details of this research and the methodology employed. Chapter IV, Results and Analysis, present the findings and the premise behind determining the merit of re-sampling to split-plot designs. Chapter V, Conclusion, summarizes the work, to include recommendations for using re-sampling techniques in the analysis of split-plot designs.

II. Literature Review

2.1 Test and Evaluation

In the DOD, T&E's fundamental purpose is knowledge gathering, in order to, assist decision makers in "managing the risks involved in developing, producing, operating, and sustaining systems and capabilities. [47]" Additionally, T&E provides knowledge of system capabilities and limitations to allow for either further developmental improvements and/or optimization of system performance by the user community. Therefore, the goal of test is the identification of deficiencies, whether technical, operational and system, early in the lifecycle, so that mitigating actions can be implemented prior to the use of the system operationally. T&E of systems may include: Developmental Test and Evaluation (DT&E), Operational Test and Evaluation (OT&E), Live Fire Test and Evaluation (LFT&E), family-of-systems interoperability testing, information assurance testing, and modeling and simulation (M&S) [47]. The type and amount of testing completed is generally decided by the system program manager (PM) and will almost always be driven by cost, schedule and performance.

2.1.1 Developmental Test. Developmental Test and Evaluation (DT&E) plans and conducts tests to determine whether the system meets its technical and performance specifications. The goal of many Developmental Testers is to test the system until it breaks. Thus, within DT&E, testers try to identify the technical capabilities and limitations of system(s), identify technical risks, stress the system under test to ensure the robustness of the system, assess technical progress and maturity against the critical technical parameters as documented in the Test and Evaluation Master Plan (TEMP) and provide support, data and analytic, on whether the system is ready for IOT&E [47]. The primary focus of DT&E is to discover and learn about the system.

2.1.2 Operational Test. Operational Test and Evaluation (OT&E) determines the operational effectiveness and suitability of the system under operationally realistic conditions against threat or threat-representative forces. OT&E also assesses

impact to combat operations and provide additional information on the system's operational capabilities [47]. The primary focus of OT&E is to assess and confirm the operational capability of the system. OT&E is a crucial element of assessing whether a system is ready for full-rate production.

2.1.3 Best Practices. "Benchmarking" is a common practice in which companies compare products, services and processes against other similar organizations to determine how they measure in regards to best practices. In fact, studies have been conducted on what are considered the "best practices." A study performed by the Science Applications International Corporation (SAIC) for the Directorate of Test, Systems Engineering and Evaluation (DTSE&E), Office of the Secretary of Defense, Washington, D.C. sought to answer the fundamental question:

> What are the best practices in Test and Evaluation that are currently employed by successful enterprises to support the maturation of product design; measure the performance of the production-ready version; and verify product acceptability for the end user application? [8]

To interview successful enterprises, SAIC designed questions of industry that address certain areas:

> Why do you test? How do you test? When do you stop? What is the value added by T&E? What do you consider your T&E best practices? Why?

Practices employed by commercial enterprises were deemed "best" practices if they:

1. Added significant value to the process by which a product was created;
2. Helped create a better product in a cheaper, faster manner; or
3. Contributed in a traceable way to the success of the company.

Among the study conclusions were: Some commercial "best" practices can be applied to DOD T&E; DOD has already identified some best practices, but they need

to be communicated more effectively; and emphasis could be increased on reducing the time required for DOD test programs. The study also recommended DTSE&E take an active role in leading the implementation of "best" practices based on commercial and DOD experience. Two areas noted by the study that merit attention are:

1. Test cycle-time reduction through the use of streamlining and appropriate "fast-track" or accelerated procedures (e.g., accomplish testing more effectively/efficiently, eliminate duplicative testing), and
2. T&E process improvement.

The study includes potential avenues of T&E process improvement: explore additional ways in which rapidly emerging information technology can be used to make T&E better, faster, and cheaper; continue to scrutinize detailed test plans to ensure that testing will generate sufficient information to address the critical issues while at the same time avoiding the expenditure of time and resources on nonessential data [8].

2.1.4 NRC Study – Dynamics in Acquisition of military systems. The Panel on Statistical Methods for Testing and Evaluating Defensive Systems was entrusted with examining the statistical techniques currently used in design and evaluation of operational tests (and can be applied to all DOD testing) in DOD and making recommendations for improvement. Cohen et al. [7] state that the acquisition of military systems is quite dynamic. Because of the dynamics, they conclude that the DOD must re-think how tests are designed, systems are evaluated and how the acquisition process is structured. They highlight five areas in which changes are occurring and challenging T&E:

1. Decreased Testing Budgets – test efforts are often smaller, shorter and have fewer prototypes;

2. More Complicated Systems – system complexity implies more measures of performance and effectiveness, which increases test design and test evaluation complexity;
3. More Software Intensive Systems – new systems require latest techniques in software engineering;
4. More Upgrades to Systems, "Evolutionary Procurement" – require the use of archived information; and
5. Greater Interest in System Reliability, Availability, and Maintainability.

Even with the decreasing test budgets that often lead to smaller and shorter tests executed with fewer prototypes, Cohen et al. conclude that more sophisticated statistical methods can help make the most effective use of whatever resources are available. In fact, they believe "even modest improvements in testing by use of the most appropriate statistical methods can lead to more efficient use of public funds and considerable improvements in the reliability and effectiveness of the systems deployed." They also assert, when appropriate, methods for combining test data with information from other sources can be used to provide additional information for decision making [7].

In essence, the advancement of technology creates more complex systems, thereby increasing the complexity of the test design and evaluation, which may require more sophisticated statistical analysis. The tests performed must produce results that permit the best decisions be made about the system. Specific techniques in experimental design have been developed to support this. These techniques are used to design tests to either maximize the information gained given a pre-specified cost or to minimize costs while providing enough information that permits a decision with acceptably small risk. Furthermore, a few experimental design principles can be applied to a wide variety of testing problems. Two particular principles are: test more where variation is expected to be the greatest and select factor levels that can best characterize

the system. Cohen at al. highlight two key problems from their examination of test designs within DOD.

> First, there is no evidence of a methodical approach to test planning, which is an important prerequisite to successful test design in industrial applications. Second, although we found many examples of the proper use of specific techniques of experimental design – including simple ideas, such as the benefits of randomization and control, and some more sophisticated designs such as fractional factorial designs – there were also test designs that were clearly not representative of the state of the art. [7]

2.1.5 Types of Tests. There are three general categories of tests:

1. Test to specification – Hypothesis test, Estimation test, Sampling plans, Quality Assurance
2. Test for problems – Intuition and experience, Edge of the envelope, Corner of the envelope
3. Test to characterize – Experimental Design

The focus of a test to specification involves using criteria that help the tester determine the merit of the system under test. The tester determines whether to pass or fail (Go/No-Go; Meets/Does not meet) the system by comparing the system against some threshold(s). Most of the time these specific thresholds are considered the Critical (Key) Performance Parameters (C(K)PP). Historically, this type of testing has been the "bread and butter" of testing. Typically, a hypothesis test is formulated. The tester will identify a test statistic used to assess the truth of the null hypothesis. After a test, a p-value is computed. The p-value is the probability that a test statistic is at least as significant as the one observed assuming that the null hypothesis were true. Many times, this type of test is conducted in a one-factor-at-a-time approach. This approach fails to consider any possible interaction between factors. Generally, a test to specification invokes α, β, power, error and sample size issues.

The focus of tests for problems involves designing the test to maximize the number of problems found at some least cost and in the shortest amount of test time.

Such tests often involve a lot of intuition and experience and could be effective if all the right subject matter experts are involved. However, this test method is poor in computing metrics or statistics, since its purpose is only to find problems. Tests designed in this fashion usually involve looking at the performance of the system in conditions generally defined as the edge of the performance envelope or the corner of the performance envelope. An assumption in this type of test is that if the system works at the edges/corners of performance, then the system will work anywhere. This is not always a valid assumption; only the behavior at the edges or corners are known when such tests are conducted.

In the test to characterize, the tester is trying to characterize the performance of the system across a variety of conditions. Tests to characterize are generally effective in finding problems and addressing issues that are inherent with the test to specification. In addition, tests to characterize are normally conducted according to a well-designed strategy (experimental design). The strategy calls for the manipulation of factors of interest in a systematic format to draw specific inferences about the effect of the factors. The objective of the experiment may include the following: [33]

1. Determine which variables are most influential on the response;
2. Determine where to set the influential factors so that the responses are almost always near the desired nominal value;
3. Determine where to set the factors so that the variability in the response is small; and
4. Determine where to set the influential factors so that the effects of the uncontrollable factors are minimized.

2.2 Design of Experiments

2.2.1 History. Montgomery discusses four eras in the modern development of statistical experimental design. The four eras include the agricultural era led by the pioneering work of Sir Ronald A. Fisher; the industrial era, catalyzed by the

development of the response surface methodology by Box and Wilson; the "quality improvement" era led by the work of Genichi Taguchi, and others; the present era led by a renewed general interest in statistical design by both researchers and practitioners [33].

During the 1920s and early 1930s, Fisher developed new methodologies for agricultural experimentation, generally regarded as the pioneering work in experimental design. He noted that agricultural experiments tend to be large and require a long time to complete. Therefore, the experimenter has to take into account for variation in the agricultural plots. He then recognized that flaws in the conduct of experiments often impacted the analysis of the data within the experiment. This recognition led to the introduction of the principles of randomization, replication, blocking, orthogonality, and statistical thinking and principles into designing experiments [33].

Box and Wilson catalyzed the industrial era with their development of response surface methodology. They recognized that industrial experiments are different from agricultural experiments based on their application of experimental design techniques to problems in the chemical industry. They noticed that they can observe a response almost immediately (immediacy), gain information from an experiment, and then apply any lessons learned to the design of the next experiment (sequentiality) [33].

In the 1970s interest in quality improvement within industry increased. This led to the "quality improvement" era. Taguchi, and others, during that era had a significant impact in the interest and use of experimental design through designed experiments. In particular, Taguchi advocated the idea of robust parameter design to improve a system or process. His intentions were to make processes less sensitive to hard to control factors (e.g. environmental factors), make products less sensitive to component variation, as well as, find levels of the process variables that tend to optimize to a desired value while also reducing the variability. His work was controversial, but had a number of positive outcomes as noted by Montgomery; "Designed experiments became more widely used in discrete parts industries and many other

industries that had previously made little use of the technique." It also help lead to the beginning of the fourth era of statistical design, in addition to the introduction of formal education in statistical experimental design in many universities [33].

The fourth era of statistical design included a renewed interest in experimental design and developed new techniques to experimental problems, including alternatives to Taguchi's methods and computer generated designs [33].

2.2.2 Strategy of Experimentation. The strategy of experimentation is a general approach to planning and conducting an experiment. In planning and conducting an experiment, an experimenter can use several strategies. Examples may include the best-guess approach, one-factor-at-a-time (OFAT) approach and specialized designs to include factorial experiments. Montgomery [33] highlights these three strategies by using a very simplistic example, golfing, and what influence four different factors had on his golf score. The four factors are:

1. The type of driver used,
2. The type of ball used,
3. Walking and carrying the golf clubs versus riding in a cart, and
4. Drinking water versus drinking beer while playing.

The best-guess approach involves selecting an arbitrary, but rationalized , combination of the factors, play golf and see what happens. During the round, it may be noticed that depending on the type of driver, the shot was impacted (several wayward shots). The next round it is decided to not use the driver that caused the wayward shots. This continues for as many rounds as played, switching the level of a factor based on the outcome of the previous round. This approach gives no guarantee of finding the best solution. Another strategy is the OFAT approach. This approach involves selecting levels, for each factor, to create a baseline. Factor levels are changed, one factor at a time, with the other factors held constant at the baseline levels. After each factor has been tested at every level, graphs are usually constructed showing the

effect the change of a level had on the response variable. The optimal combination is selected from the graphs. OFAT experiments fail to consider any possible interactions between the factors and are less efficient than other methods based on a statistical approach to design.

The final strategy discussed is the factorial experiment. Here, the factors are varied together, instead of one at a time. With this type of experiment the experimenter can investigate the individual effects of each factor and consider any possible interactions that exist between factors. An advantage of this approach is that it makes the most efficient use of the experimental data. A factorial experiment is a specific, and very popular design within the Design of Experiments (DOE) paradigm.

2.2.3 What is Design of Experiments? DOE is a systematic and rigorous process of planning, conducting and analyzing experiments, a specialized form of experimental design. It involves planning the experiment so that appropriate data is collected and analyzed using proper statistical methods. This approach seeks valid and objective conclusions [33]. A poor design may capture little information, so great thought by the experimenter working with subject matter experts is needed. Designing an experiment means taking the time and effort to properly organize the experiment to ensure that the correct data is available (type and amount) to formalize the conclusions as clearly and efficiently as possible. The primary goal of an experimental design of this type is to establish (or rule out) a cause-effect relationship between the independent and dependent variables [33]. In addition, DOE is meant to extract the maximum amount of information with minimal cost.

The basic principles of DOE are randomization, replication, blocking and orthogonality. Randomization implies that individual experiments are performed in random order. Randomization has three purposes. First, randomization helps to evenly distribute system or process idiosyncratic characteristics, so as to not bias the outcome of the experiment. Second, randomization allows the computation of an unbiased estimate of error effects. Third, randomization helps to ensure that the

error effects are statistically independent, a requirement for many statistical methods [29] [5].

A replication is an independent repeat of a factor combination comprising an individual experiment. Replication provides an unbiases estimate of true experimental error. This estimate becomes a key measurement in determining statistical differences in the data. The more replications, the better the estimate of the experimental error, and the more precise the estimate of the response of interest. Replication reflects sources of variability within runs.

Blocking is an experimental procedure used to improve precision with which comparisons among the factors are made. Blocking is used to isolate the variation attributed to a nuisance factor. The nuisance factors are factors that may influence the response that are uncontrollable and blocking attempts to reduce or eliminate that variance. The orthogonality of the test conditions immediately implies that all the test conditions are inherently independent [33]. Therefore, a design that is orthogonal is advantageous.

Common experimental designs include: [33]

1. 2^k factorial design – A design that has k factors, each at only two levels ("high" and "low"), particularly useful in the early stages of experimental work when many factors are likely to be investigated, widely used in factor screening experiments and in sequential experimentation.
2. 3^k factorial design – A factorial arrangement with k factors, each at three levels ("low", "intermediate" and "high"), allows for a quadratic relationship between response and design factors.
3. Mixed-level factorial design – Factors have varied levels, mostly two or three, and usually occur when there are both quantitative and qualitative (mixed) factors in the experiment.
4. Fractional factorial design – These designs are used in screening experiments used to try and reduce a large set of experimental factors down to a smaller,

more manageable set. Their success is based on three key ideas: Sparsity of effects principle (system is driven by main effects and low-order interactions), the projection property (design projected into stronger designs in the subset of significant factors) and sequential experimentation (combine runs from other fractional factorials to assemble a larger design).

5. Response surface design – Designs for first-order and second-order models, most often used to build models for making predictions, determining optimality, and characterizing system surfaces that are non-linear in structure.
6. Nested designs – Levels of one factor are similar but not identical for different levels of another factor,
7. Split-plot designs – These designs are used when it is impossible to run a CRD due to limitations involving time, material, cost and resources. These designs typically fix the levels of hard-to-change (HTC) factor and run all combinations of the other factors for each HTC factor setting.

2.2.4 Why DOE? DOE is a multipurpose methodology that can be used in many situations. A test designed using the principles of DOE yields a more effective method of test. The structure of DOE allows an experimenter to gain more insight faster and at a lower cost. Fewer runs are typically needed for a test conducted in a purposeful manner, and at the same time DOE provides information about the interaction of factors and the way the total system works, something that cannot be understood OFAT testing.

DOE provides other benefits to a test. DOE can avoid the confounding of effects that may occur when there is not a systematic approach to the design and conduct of the test. DOE can also help determine the important variables that need to be controlled and at the same time help determine the unimportant variables that may not need to be controlled.

Additional advantages of DOE include: [45]

1. DOE provides a structured planning process used to involve stakeholders to generate test and analysis plans that are comprehensive and efficient;
2. Sequential testing and analysis leads to quicker system discovery and understanding; and
3. Empirical statistical models can be used for estimation and prediction of system response functions.

2.2.5 What Applications? DOE has been applied in many functional areas including Research ([25]), Product development ([28], [23]), Quality Control ([34], [49]), Market Research ([42], [48]), and Engineering. In research, DOE has been used to quantify interrelationships between variables and to screen large sets of variables to find important subsets of variables. Product development has used DOE to improve products through reformulation and improvement of products as well as in development of new products. DOE is used in quality control in setting specifications on quality characteristics. In addition, DOE is used in market research to measure consumer preference for products and determine how to optimize the sale of products among consumers.

Other specific examples of DOE include process characterization, process validation, process optimization, simulation, robust parameter design and Military T&E [6].

2.2.6 What is a Split-Plot Design? A specialized design that has been used by experimenters is a split-plot design. A split-plot design is a multifactor factorial experiment in which the experimenter is unable (or doesn't choose) to completely randomize the order of the runs in at least one of the factors in the design.

Split-plot designs have three main characteristics: [24]

1. The levels of all the factors are not randomly determined and reset for each experimental run – A HTC factor is held at a particular setting and all combinations of the other factors are run.

2. The size of the experimental unit is not the same for all factors – A factor is applied to a larger group involving combinations of the other factors; whole plot versus subplot.

3. There is a restriction on the random assignment of the treatment combinations to the experimental units – A prohibition in assigning treatments to the units completely randomly.

The most frequently encountered situations where split-plot designs occur are:

1. When an experiment consists of two types of experimental units – some factors require large experimental units (whole plots) and others require small experimental units (subplots), or

2. When some factor levels are easy or inexpensive to change (ETC) while others are HTC – HTC factors form the whole plots; ETC factors form the subplots.

2.2.7 Split-Plot Design Model Examples. A great example of a split-plot design can be found in agricultural research, where it is common to experiment on plots (fields) of land. For example, several varieties of a crop are planted in different fields. Each field is divided into multiple subplots and each subplot is treated with a different type of fertilizer. In this case, the different crops represent the main treatments (whole-plot) and the different fertilizers are the sub-treatments (subplots).

The linear model for the split-plot design, when considering two factors, is the following:

$$y_{\mathbf{ijk}} = \mu + \tau_{\mathbf{i}} + \beta_{\mathbf{j}} + (\tau\beta)_{\mathbf{ij}} + \gamma_{\mathbf{k}} + (\tau\gamma)_{\mathbf{ik}} + (\beta\gamma)_{\mathbf{jk}} + (\tau\beta\gamma)_{\mathbf{ijk}} + \epsilon_{\mathbf{ijk}} \tag{2.1}$$

where τ corresponds to the effects represented by blocks or replicates, β corresponds to the effects due to main treatments (factor A),$\tau\beta$ corresponds to the whole-plot error, γ corresponds to the subplot treatment (factor B), $\tau\gamma$ corresponds to the block or replicate interaction with B, $\beta\gamma$ corresponds to the interaction between factors A

and B, and $\tau\beta\gamma$ is the subplot error. [33] An alternative form of the model above is the following:

$$y_{\mathbf{ijk}} = \mu + \tau_{\mathbf{i}} + \beta_{\mathbf{j}} + (\tau\beta)_{\mathbf{ij}} + \gamma_{\mathbf{k}} + (\tau\gamma)_{\mathbf{ik}} + (\beta\gamma)_{\mathbf{jk}} + \epsilon_{\mathbf{ijk}} \tag{2.2}$$

$(\tau\beta)_{\mathbf{ij}}$ is still the whole-plot error, however $\epsilon_{\mathbf{ijk}}$ now represents the subplot error. If it is reasonable to assume that the replicate and Factor B interaction, along with, replicates and Factor A, Factor B interaction are negligible then this alternative is satisfactory. [33]

The model expands when additional factors are added. For example, consider an experiment with four design factors (A, B, C, D). Now factors A and B are difficult to change, whereas C and D are easy to change. The model for this experiment is the following:

$$y_{\mathbf{ijklm}} = \mu + \tau_{\mathbf{i}} + \beta_{\mathbf{j}} + \gamma_{\mathbf{k}} + (\beta\gamma)_{\mathbf{jk}} + \theta_{\mathbf{ijk}} + \delta_{\mathbf{l}} + \lambda_{\mathbf{m}} + (\delta\lambda)_{\mathbf{lm}} + (\beta\delta)_{\mathbf{jl}} + (\beta\lambda)_{\mathbf{jm}} + (\gamma\delta)_{\mathbf{kl}} +$$

$$(\delta\lambda)_{\mathbf{lm}} + (\beta\gamma\delta)_{\mathbf{jkl}} + (\beta\gamma\lambda)_{\mathbf{jkm}} + (\beta\delta\lambda)_{\mathbf{jlm}} + (\gamma\delta\lambda)_{\mathbf{klm}} + (\beta\gamma\delta\lambda)j_{\mathbf{klm}} + \epsilon_{\mathbf{ijklm}}$$

τ represents the replicate effect, β and γ represents the whole plot main effects, θ is the whole plot error, δ and λ represent the subplot main effects, and ϵ is the subplot error. [33]

2.2.8 Split-Plot Analysis. The analysis of a split-plot experiment is easiest if done with two separate analyses, one for the whole plot and the other for the subplot. As is typical with other experimental designs, the null hypothesis, H_0, is that there is no effect due to a factor. However, since the analysis is performed first for the whole plot and then for the subplot, different criteria are used in forming the associated test F-statistics. In particular, the F-statistic is the ratio between the mean square of the

Table 2.1: General ANOVA for Split-Plot Analysis

Sources of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Replicates	$SS_{replicate}$	$r-1$	$\frac{SS_{replicate}}{r-1}$	
Factor A	SS_A	$a-1$	$\frac{SS_A}{a-1}$	$\frac{MS_A}{MS_{WPerror}}$
Whole Plot Error	$SS_{WPerror}$	$(r-1)(a-1)$	$\frac{SS_{WPerror}}{(r-1)(a-1)}$	
Factor B	SS_B	$b-1$	$\frac{SS_B}{b-1}$	$\frac{MS_B}{MS_{SPerror}}$
Factor AB	SS_{AB}	$(a-1)(b-1)$	$\frac{SS_{AB}}{(a-1)(b-1)}$	$\frac{MS_{AB}}{MS_{SPerror}}$
Subplot Error	$SS_{SPerror}$	$a(r-1)(b-1)$	$\frac{SS_{SPerror}}{a(r-1)(b-1)}$	
Total	SS_{Total}	$rab-1$		

factor of interest to the correct mean square error component.

$$F = \frac{MS_{factor}}{MS_{correcterror}} \tag{2.3}$$

In the case of the whole plot factor(s), the mean square error is the mean square error of the whole plots, $MS_{WPerror}$. The mean square error for the subplot factor(s) is the mean square error for the subplot, $MS_{SPerror}$.

Table 2.1, summarizes the analysis of variance (ANOVA) for the case represented in equation 2.2 where there are only two factors, a whole plot factor (a levels) and a subplot factor (b levels).

The following example comes from Montgomery [33]. A two factor, split-plot design involves a paper manufacturer who is interested in three different pulp preparation methods (Factor A) and four different cooking temperatures for the pulp (Factor B). The manufacturer wants to study the effect these two factors have on the overall tensile strength of the paper. In this case, Factor A is the whole plot factor and Factor B is the subplot factor. Twelve observations are required to complete each replicate of the factorial experiment and three replicates are needed. However, only 12 runs are capable in a day. The experiment is then conducted, such that, a batch

Table 2.2: Experiment on the Tensile Strength of Paper from Montgomery (2007)

Pulp	**Replicate 1**			**Replicate 2**			**Replicate 3**		
Preparation Method	1	2	3	1	2	3	1	2	3
Temperature (F)									
200	30	34	29	28	31	31	31	35	32
225	35	41	26	32	36	30	37	40	34
250	37	38	33	40	42	32	41	39	39
275	36	42	36	41	40	40	40	44	45

Table 2.3: ANOVA of Tensile Strength of Paper example

Sources of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0
Replicates	77.556	2	38.778	
Factor A	128.389	2	64.194	7.08
Whole Plot Error	36.278	4	9.069	
Factor B	434.083	3	144.694	36.43
Factor AB	75.167	6	12.528	3.15
Subplot Error	71.500	18	3.972	
Total	822.972	35		

of pulp is prepared by one method, split into four samples and observations for all four temperatures are obtained from that batch. A total of 36 observations are made with 9 different batches. This is a split-plot design and the analysis performed is a split-plot analysis. Table 2.2 and Table 2.3 summarizes the data for the experiment and analysis on the tensile strength of paper, respectively.

Initially, the whole plot analysis is conducted. In the whole plot analysis, the source of variation that is of interest is replicates (or blocks), pulp preparation method (Factor A) and the whole plot error.

The sums of squares are computed as follows:

$$SS_{replicate} = \sum \frac{Y_{i..}^2}{ab} - \frac{Y_{...}^2}{abr}$$

$$= \frac{(417^2+423^2+457^2)}{12} - \frac{1297^2}{36}$$

$$= 77.556$$

$$SS_{FactorA} = \sum \frac{Y_{.j.}^2}{br} - \frac{Y_{...}^2}{abr}$$

$= \frac{(428^2+462^2+407^2)}{12} - \frac{1297^2}{36}$

$= 128.39$

$SS_{WPerror} = SS_{WP} - SS_{FactorA} - SS_{replicate}$

$= \sum \frac{Y_{ij.}^2}{b} - \frac{Y_{...}^2}{abr} - 128.39 - 77.556$

$= \frac{(138^2+155^2+124^2+141^2+149^2+133^2+149^2+158^2+150^2)}{4} - \frac{1297^2}{36} - 128.39 - 77.556$

$= 36.276$

The mean squares are computed as follows:

$MS_{replicate} = \frac{SS_{replicate}}{r-1}$

$= \frac{77.556}{2}$

$= 38.778$

$MS_{FactorA} = \frac{SS_{FactorA}}{a-1}$

$= \frac{128.39}{2} = 64.195$

$MS_{WPerror} = \frac{SS_{WPerror}}{(r-1)(a-1)} = \frac{36.276}{4} = 9.069$

The F-statistic for Factor A is computed as follows:

$F_{FactorA} = \frac{MS_{FactorA}}{MS_{WPerror}} = \frac{64.195}{9.069} = 7.0785$

Finally, the subplot analysis is conducted. In the subplot analysis, the source of variation that is of interest is temperature (Factor B), AB interaction and the subplot error.

The sums of squares are computed as follows:

$SS_{FactorB} = \sum \frac{Y_{..k}^2}{ar} - \frac{Y_{...}^2}{abr} = \frac{(281^2+311^2+341^2+364^2)}{9} - \frac{1297^2}{36} = 434.08$

$SS_{AB} = \sum \frac{Y_{.jk}^2}{r} - \frac{Y_{...}^2}{abr} - SS_{FactorA} - SS_{FactorB}$

$= \frac{89^2+100^2+92^2+104^2+117^2+90^2+118^2+119^2+104^2+117^2+126^2+121^2}{3} - \frac{1297^2}{36} - 562.47$

$= 637.64 - 562.47$

$= 75.17$

$SS_{SPerror} = SS_{Total} - SS_{replicate} - SS_{FactorA} - SS_{WPerror} - SS_{FactorB} - SS_{AB}$

$= \sum Y_{ijk}^2 - \frac{Y_{...}^2}{abr} - 77.556 - 128.39 - 36.276 - 434.08 - 75.17$

$= 822.97 - 751.47$

$= 71.5$

The mean squares are computed as follows:

$MS_{FactorB} = \frac{SS_{FactorB}}{b-1} = \frac{434.08}{3} = 144.69$

$MS_{AB} = \frac{SS_{AB}}{(a-1)(b-1)} = \frac{75.17}{6} = 12.528$

$MS_{SPerror} = \frac{SS_{SPerror}}{a(r-1)(b-1)} = \frac{71.5}{18} = 3.9722$

The F-statistics for Factor B and the AB interaction are computed as follows:

$F_{FactorB} = \frac{MS_{FactorB}}{MS_{SPerror}} = \frac{144.69}{3.9722} = 36.426$

$F_{AB} = \frac{MS_{AB}}{MS_{SPerror}} = \frac{12.528}{3.9722} = 3.1539$

2.2.9 Split-plot Advantages and Disadvantages. Advantages of a split-plot design include:

1. It provides an efficient use of factors requiring large experimental units in combination with other factors requiring small experimental units; it allows them to be tested in the same experiment [38].
2. It allows increased precision for comparing certain factors, as compared to a Randomized Block Design. Subplot variance is generally less than whole-plot variance, the subplot treatment factor and interaction are generally tested with greater sensitivity [38].
3. It allows the introduction of new treatments into an experiment already in progress. A second factor may be included at very little cost [38].
4. It can combine experiments in which some factors require large amounts of experimental material and other factors require very little material [39].
5. It is a natural way to handle repeated measurements [39].

6. It helps in saving experimental material [39].

Disadvantages of a split-plot design:

1. Analysis is complicated by the presence of two experimental error variance components [38].
2. Low precision on the whole plot errors can result in large differences being insignificant, while small differences on the subplots may be statistically significant even though they are of no practical significance [38].
3. When missing data occur, the analysis is typically more complicated than for a randomized complete block design with missing data [39].

In order to compensate for the differences in size and precision of the whole plot and subplot factors, the following are considered:

1. If more precision is needed for some factor B compared to another factor A, assign factor B to the subplots and factor A to the whole plots
2. If the main effect of one factor (factor A) is expected to be much larger, and easier to detect as significant, than that of the other factor (factor B), factor A should be assigned to the whole plots and factor B to the subplots.
3. If experimental practices require a factor to use large plots, assign that factor to whole plots.

2.3 Re-sampling

Prior to the advancements in computer processing many researchers embraced traditional statistical methods rather than experimenting with new techniques, such as re-sampling methods. Three factors contributed to this practice. First, new methods were not readily known and the concepts tended to remain obscure. Textbooks do not include the advanced techniques immediately; there is typically a time delay for inclusion. Second, software programs previously were devoted to conventional data

analysis and did not always include the new techniques. Even if a researcher was aware of newer techniques, the limited software availability led to the use of traditional methods. Third, traditional procedures are perceived as founded on solid theoretical and empirical justification, while new techniques face initial criticism and may lack accepted justification [52].

The continued use of traditional methods over newer methods does come with a price. Only certain types of statistics are analyzed, such as the mean and standard deviation. In addition, certain assumptions about the underlying data distribution are usually needed, like the normality assumption. Finally, researchers need specialized training to apply, understand, and appreciate statistics [40].

Today, with the advancements in computer processing, re-sampling methods are aided by high-speed computers since all techniques rely on the computer to generate data sets from the original data. Thus, re-sampling methods have become increasingly popular as statistical tools. They have overcome the limitations presented previously. Virtually any statistic can be analyzed, no assumptions are needed about the distribution of the data and the techniques are easily understood. Also, the methods are very robust, and their computational demands are no longer an issue.

2.3.1 What is re-sampling? Re-sampling refers to a variety of statistical methods based on available data rather than on a set of assumptions about the underlying population. In re-sampling, the basic idea is to mimic the process of sampling by picking samples at random from a hypothetical population of interest, based on a sample from that population, to draw improved inferences about the population. Usually, in order to draw inferences, many samples are needed from the population. At times it becomes too expensive or impractical to sample more data from the population itself. Instead, sample variability is studied using re-sampling methods constructed on the computer (Monte Carlo simulations). However, no more information is provided about the population other than that obtained from the original

sample data using these methods, but it can provide a way to draw inferences about the population based on the sampled data set where traditional methods could not.

2.3.2 What are some re-sampling methods. Re-sampling methods include permutation tests, jackknife methods, cross-validation, and bootstrap methods. They are used to perform many functions to include:

1. Estimating the precision of the sample statistics by using subsets of available data.
2. Estimating the precision of the sample statistics by drawing randomly with replacement from a set of data points.
3. Exchanging labels on the data points when performing significance tests.
4. Validating models by using random subsets.

Permutation tests involve the shuffling of the observed data to determine how unusual an observed outcome is. Jackknife methods involve computing the statistic of interest for all combinations of the data where one (or more) of the original data points are removed. Cross-validation uses part(s) of the available data to fit a model and the remaining part(s) to test the model. Bootstrap methods attempt to estimate the sampling distribution of a population by generating new samples by drawing (with replacement) from the original data. Each are discussed further; bootstrap methods are the focus of this research.

2.3.3 Permutation Tests. Permutation tests are a computer-intensive statistical technique introduced by R.A. Fisher in the 1930's. The idea predates computers and was introduced more as a theoretical argument supporting Student's t-test than as a useful statistical method [19]. Modern computational power makes permutation tests practical to use. The permutation test is a non-parametric test and requires no particular assumptions concerning statistical distributions, they are increasingly applied even in the context of traditional tests such as correlation, t-tests, and ANOVAS.

A typical permutation test problem involves testing the hypothesis that two or more samples might belong to the same population. The test proceeds as follows:

1. Obtain observational samples.
2. Devise a test statistic, θ.
3. Calculate test statistic on the obtained data, $\hat{\theta}_{original}$.
4. Define a null hypothesis,H_0.
5. Randomly rearrange data to create permutation sample.
6. Calculate test statistic for permutation sample, $\hat{\theta}_n$ where $n = 1, 2, \ldots, N$. Record the statistic of interest.
7. Repeat Steps 5-6 N times, such that N is a large number to create empirical distribution of the test statistic.
8. Compare $\theta_{original}$ to empirical test statistic distribution. If true test statistic is greater than $(1 - \alpha)$ percent of the random values, then the null hypothesis is rejected at $p < \alpha$.

   Further information on permutation tests is included in [19], [20], [37].

2.3.4 Jackknife. The jackknife method introduced by Quenouille, and further developed by Tukey, is a technique for estimating the bias and standard error of an estimate. [32] The jackknife is less dependent on model assumptions and does not need the theoretical formula required by the traditional approaches. However, it does require computing the statistic m times, therefore prior to the advancements in computer processing it was not a popular method.

The jackknife provides a way of decreasing bias and obtain standard errors in situations where the standard methods might be inappropriate (i.e., distribution of the sample is not normal). The jackknife method works by calculating the statistic (or statistics) of interest, omitting each data value in turn. These "partial estimates" are then combined with the estimate obtained from the inclusion of all sample points

to produce the pseudo-values. The jackknife estimate of the statistic involves the mean and standard error of the pseudo-values.

The jackknife proceeds as follows:

1. Obtain an observational sample, $\mathbf{X} = (X_1, \cdots, X_m)$.
2. Devise point estimate, θ.
3. Calculate point estimate on the obtained data, $\hat{\theta}_{original}$.
4. Create jackknife samples that leave out j^{th} observation

$$\mathbf{X}_{(j)} = (X_1, X_2, \cdots, X_{j-1}, X_{j+1}, \cdots, X_m).$$

   for $j = 1, 2, \cdots, m$.

5. Calculate point estimate on jacknife samples, $|hat\theta_{(j)}$.
6. Calculate pseudovalue on jackknife samples

$$\mathbf{p}_j = m\theta_{original} - (m-1)\hat{\theta}_{(j)}.$$

   for $j = 1, 2, \cdots, m$.

7. Calculate jackknife estimate for the point estimate, $\hat{\theta}$.

$$\mathbf{p} = \frac{\sum \mathbf{p}_j}{m}.$$

   for $j = 1, 2, \cdots, m$.

Further information on jackknife methods is included in [31], [32], [19], [14], [17].

2.3.5 Cross-validation. Prediction error measures how well a model predicts the response value of a future observation. Since it is sensible to choose a model that has the lowest prediction error among a set of candidates, it is often used for model selection. Cross-validation is a method used for estimating prediction error [14].

Usually, data is limited because of insufficient resources. Cross-validation uses part of the available data to fit the model, and a different part to test it. When there are large amounts of data, the data are commonly split into two equal parts. When there is not, K-fold cross-validation is used to make more efficient use of the the available information. K-fold cross-validation proceeds as follows [14]:

1. Split the data into K roughly equal-sized parts.
2. For the k^{th} part, fit the model to the other $K-1$ parts of the data, and calculate the prediction error of the fitted model when predicting the k^{th} part of the data.
3. Do the above for $k = 1, 2, \cdots, K$ and combine the K estimates of prediction error.

Further information on cross-validation is included in [44], [16], [19].

2.3.6 Bootstrap. The bootstrap method was introduced by Efron [14] as a computer-based method for estimating the standard error of the point estimate and is described in depth in Efron and Tibshirani [19]. Additional resources that helped in the generation of the information below on bootstrap include: [18], [50], [43], [51]. The idea behind the bootstrap is that in the absence of any other knowledge about a population, the distribution of values found in a random sample of size m best represents the distribution in the population. The bootstrap uses the original population sample and increases the sample size by re-sampling from that population. A benefit of the bootstrap methodology is that it requires no theoretical calculations, and can be found no matter how complicated the point estimator may be.

A bootstrap sample $X^* = (X_1^*, X_2^*, \cdots, X_m^*)$, is obtained by randomly sampling m times, with replacement, from the original observational sample ($X = (X_1, X_2, \cdots, X_m)$). For each independent bootstrap sample, $X^{*1}, X^{*2}, \cdots, X^{*B}$ (B is the number of bootstrap samples generated), a bootstrap replication (the value of the statistic of interest for the bootstrap samples) is calculated. Figure 2.1 is a schematic of the bootstrap process.

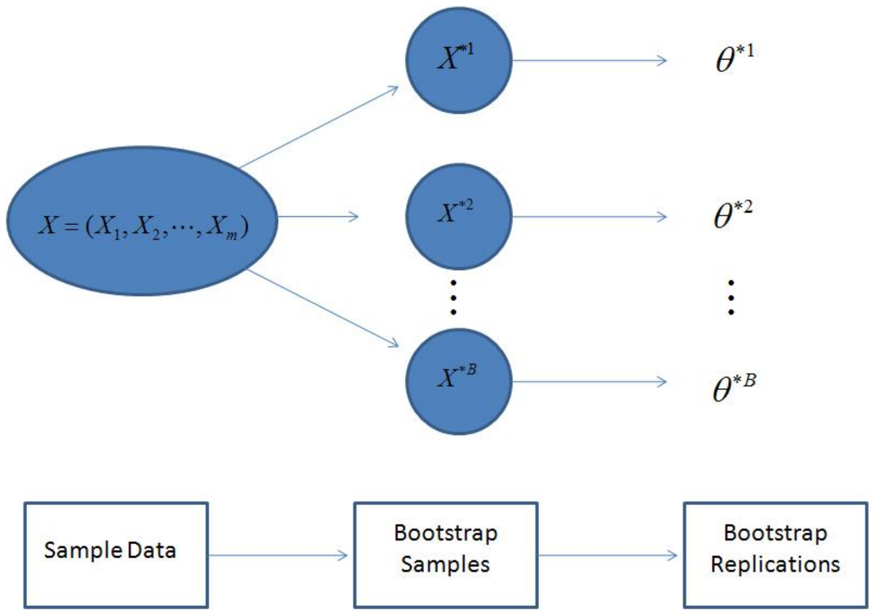


Figure 2.1: Bootstrap Process Schematic – General.

The bootstrap has two important assumptions:

1. The original sample is a valid representative of the population.
2. Each observation in a sample is independent and identically distributed (i.i.d).

Advantages of the bootstrap include:

1. The bootstrap is quite general;
2. It is a nonparametric approach and does not require distributional assumptions; and
3. Users can apply bootstrap to statistics with sampling distributions that are difficult to derive.

Disadvantages of the bootstrap include:

1. Bootstrap is sensitive to outliers in the data set; and
2. It is a computer intensive method.

The general procedure for the bootstrap is as follows:

1. Obtain an observational sample, $(X_1, \cdots, X_m)$.
2. Draw B independent bootstrap samples from the original sample of size m
3. Estimate the parameter of interest for each bootstrap sample θ^{*b}, where $b = 1, 2, \cdots, B$.
4. Generate mean for the bootstrap replications. $\theta^* = \frac{1}{B}(\sum_{b=1})^B \theta^{*b}$
5. Estimate the standard error for the estimator by finding the standard deviation of the bootstrap replications.

There is no general agreement on the number of bootstrap replications needed in bootstrap. For estimating errors, B is usually 50-250, and for bootstrap confidence intervals, a much larger B is required, 500-10,000 [46].

2.3.7 Bootstrap Confidence Interval Methods. Many methods have been used to formulate the bootstrap confidence interval and include: the bootstrap percentile method, bootstrap-t methods, BC_a method, ABC method.

The bootstrap percentile method is popular due to its simplicity. After the conduct of $B = 1000$ bootstrap replications of θ^* , the bootstrap replications are rank ordered, smallest to largest. Then, the two-tailed bootstrap percentile confidence interval at 95 percent level of confidence is the 25th entry, $B_{.025}$, and the 975th entry, $B_{.975}$. These confidence intervals in general are not symmetric. The centered version of the bootstrap percentile method states that the real valued estimator θ lies within the range $(2\theta^* - B_{.975}, 2\theta^* - B_{.025})$.

The Bootstrap-t procedure estimates the t-distribution directly from the observational sample. This estimation is used as the test statistic to formulate the confidence intervals. Further information on the Bootstrap-t is found in [14], [21], [13].

BC_a method is an automatic algorithm for producing confidence intervals from a bootstrap distribution. This method relies on the cumulative distribution function (CDF) of the bootstrap replications and two numerical parameters: the bias correction

z_0 and the acceleration a. Further information for this method can be found in [15], [21], [19], [13].

The ABC method or approximate bootstrap confidence intervals method is a method of approximating the BC_a interval endpoints analytically, without using Monte Carlo replication. It works by approximating the bootstrap random sampling results using a Taylor series expansion. DiCiccio and Efron [12] introduced this method and it is discussed in [19], [13].

2.3.8 Different Bootstrap Methods. In addition to the nonparametric bootstrap described previously, other variations include the parametric bootstrap, wild bootstrap, smoothed bootstrap, m-Out-of-n-bootstrap, iterated bootstrap, balanced bootstrap and blocked bootstrap.

With the parametric bootstrap, a distribution model is fit to the data, often by maximum likelihood. Bootstrap samples are then drawn from the distribution model. The parameter of interest is computed from these samples as with the non-parametric bootstrap. Typically, assumptions are made regarding the underlying distribution of the population [19]. The wild bootstrap is generally used in a regression setting with heteroscedasticity issues. It proposes to multiply each residual independently by a random variable with expectation zero and variance one. The technique is developed in [28] and discussed in [30], [11], [10], [9]. The smoothed bootstrap is typically referred to as an intermediate solution between parametric and nonparametric bootstrapping. Instead of re-sampling directly from the empirical distribution, the distribution is smoothed first and then the smoothed version is used to generate the new samples. A simpler method adds a small amount of random noise to each bootstrap observation. Further information on the smoothed bootstrap is contained in [41], [22], and [19]. The m-Out-of-n-bootstrap is a fairly new approach with active research in the area. It appears to be a very general way to resolve bootstrap failure by forming smaller bootstrap samples from larger samples. Work is found in [2] and [1]. The Iterated bootstrap, or double bootstrap, involves bootstrapping the bootstrap sam-

ples. Discussion on this is in [11], [27], [35]. The balanced bootstrap is an alternative sampling method that forces each observation to occur a total of $m \times B$ times in the collection of $m \times B$ bootstrap samples. The balanced bootstrap is further examined in [4] and [36]. The blocked bootstrap is used in the case of dependent observations, where the ordinary bootstrap fails, since bootstrap samples are drawn independently from the original sample. A way to overcome the failure is by re-sampling blocks of consecutive observations. Particular block bootstrap methods are discussed in [26].

2.4 Relation to Methodology

DOD test resources are limited. DOD test conduct often faces randomization restrictions. Many of the resulting tests take on a split-plot structure. Re-sampling methods have been successfully applied in a variety of statistical settings. Bootstrap re-sampling may have applicability in DOD test as a basis for improving the precision associated with the error estimates in split-plot test analysis.

III. Methodology

This research examines the application of bootstrapping to potentially improve the error estimation in split-plot experiments. For various split-plot designs a theoretical model is defined and sampled to create split-plot design experimental results. These theoretical models include defined whole-plot and subplot error components. The results are then bootstrapped and analyzed to assess any improvements in error estimation.

3.1 Monte Carlo

Monte Carlo simulations are methods to iteratively evaluate deterministic models using random numbers as inputs. The idea behind Monte Carlo simulations is to use random samples of inputs to explore the dynamic behavior of a process. The Monte Carlo methodology was first employed by scientists working on nuclear weapons projects in the 1940s, as part of the Los Alamos National Laboratory. No single approach for the Monte Carlo method is used; a number of approaches exist. Monte Carlo approaches tend to have the following pattern:

1. Define the domain of possible inputs.
2. Generate inputs randomly from the domain using a specified probability distributions.
3. Perform a deterministic computation using the inputs.
4. Aggregate the results of the individual computations into the final results. In general, Monte Carlo is used to refer to any type of random sampling empirical study.

3.2 Bootstrap applied to Linear regression

Bootstrap techniques can be applied to linear regression model selection. Most of the bootstrap techniques when applied to linear regression use the ordinary least

squares (OLS) procedures to estimate the parameters of the model. In the regression setting, there are two different ways to conduct the re-sampling [46]:

1. The regressor(s) is random (Random-x re-sampling).
2. The regressor(s) is fixed (fixed-x re-sampling).

In the fixed-x re-sampling a particular method has been developed by Efron and Tibshirani called the classical bootstrap fixed-x re-sampling method (CBRM). The procedure is summarized as follows [19]:

1. Step 1: Fit the OLS to the original sample of observations to get the fitted values.
2. Step 2: Obtain the residuals.
3. Step 3: Draw n bootstrap random samples with replacement from the residuals.
4. Step 4: Fit the OLS to the bootstrapped values.
5. Step 5: Repeat steps 3 and 4 B times, where B is the bootstrap replications.

3.3 Split-Plot Data Generation

The split-plot designs represented in Table 3.1 (varying from a single whole plot and a single subplot factor to five whole plot and five subplot factors) were used to generate samples and examine the applicability of bootstrapping to improve error estimates of whole plot and subplot errors in split-plot analysis.

A defined split-plot model is used to generate the data for each Monte Carlo simulation. A different model is defined for each of the designs used. The model defined for each design includes coefficients for the intercept, the main effects and two-way interactions. In addition, the random errors are generated based on the error structure for split-plot experiments defined by Bisgaard and de Pinho [3]; they explain the two-stage split-plot randomization and why it is appropriate to use two separate normal plots for the analysis of two-level factorial split-plot experiments. (See

Table 3.1: Split-Plot Designs

Design	Whole Plot Factors	Subplot Factors
1	1	1
2	1	2
3	1	3
4	2	1
5	2	2
6	2	3
7	3	1
8	3	2
9	3	3

Figure 3.1). The hierarchical structure of a two-level, split-plot experiment involves a random error, $\varepsilon_{\mathbf{i}}$, with standard deviation, σ_1,between whole-plot trials and another random error, $\varepsilon_{\mathbf{ij}}$, with standard deviation, σ_0, between subplots. The error structure, defined above and illustrated further in Figure 3.2, acknowledges that subplot trials within the same whole-plot are more alike than subplot trials from different whole-plots. Thus, the combined error for each observation,or trial, is the sum of the two errors.

$$E_{\mathbf{k}} = \varepsilon_{\mathbf{i}} + \varepsilon_{\mathbf{ij}} \tag{3.1}$$

When conducting a Monte Carlo simulation to study split-plot design analysis, the two random errors that represent the combined error, Equation 3.1, are assigned via random number draws from two normal distributions representing the two distinct errors, both with mean of zero and standard deviation of σ_1 and σ_0, respectively (i.e., $\varepsilon_{\mathbf{i}} \sim Norm(0, \sigma_1)$ and $\varepsilon_{\mathbf{ij}} \sim Norm(0, \sigma_0)$). A random draw is performed for each distinct whole plot and subplot (ar and abr random draws needed respectively; a represents the number of whole plots in a single replication of the design; b represents the number of subplots within each whole plot; r represents the number of replications observed). For the study, 13 sets of distributions were used to represent the respective errors. The sets are included in Table 3.2.

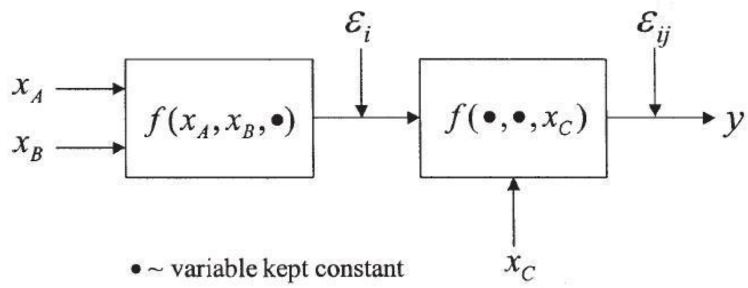
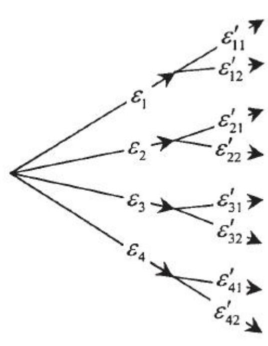


Figure 3.1: Two-stage randomization in a split-plot experiment from Bisgaard (2004). Two errors: the whole plot error $\varepsilon_{\mathbf{i}}$ and subplot error $\varepsilon_{\mathbf{ij}}$.



	A	B	C	AB	AC	BC	ABC
$\varepsilon_1+\varepsilon'_{11}$	−	−	−	+	+	+	−
$\varepsilon_1+\varepsilon'_{12}$	−	−	+	+	−	−	+
$\varepsilon_2+\varepsilon'_{21}$	+	−	−	−	−	+	+
$\varepsilon_2+\varepsilon'_{22}$	+	−	+	−	+	−	−
$\varepsilon_3+\varepsilon'_{31}$	−	+	−	−	+	−	+
$\varepsilon_3+\varepsilon'_{32}$	−	+	+	−	−	+	−
$\varepsilon_4+\varepsilon'_{41}$	+	+	−	+	−	−	−
$\varepsilon_4+\varepsilon'_{42}$	+	+	+	+	+	+	+

Figure 3.2: Error Structure of 2^3 factorial experiment from Bisgaard (2004). A and B are whole plot factors and C is subplot factor.

Table 3.2: Standard Deviations for Error Distribution Sets

Distribution Structure	Whole Plot Error σ_1	Subplot Error σ_0
1	2	2
2	2	4
3	2	6
4	2	8
5	2	10
6	4	4
7	6	6
8	8	8
9	10	10
10	4	2
11	6	2
12	8	2
13	10	2

The observation data for the designs indicated in Table 3.1, is then represented by the following:

$$Y_{\mathbf{k}} = X_{\mathbf{k}} * C + E_{\mathbf{k}}, \tag{3.2}$$

such that $Y_{\mathbf{k}}$ represents the k^{th} observation generated, C represents the coefficients for the design model, $X_{\mathbf{k}}$ represents the k^{th} augmented design point (Augmented design point includes a column to represent the intercept, each factor and two-way interaction), and $E_{\mathbf{k}}$ is the combined random error for the k^{th} design point.

3.4 Example Analysis

An example for three replications of Design 1 indicated in Table 3.1 is presented. Equation 3.3 is the theoretical model used in the simulation for Design 1.

$$E(y) = 50 + 10A + 5B + 2AB \tag{3.3}$$

where A, B, are the setting levels for factor A, factor B, respectively.

1. Define X:

   For this example, there is only one whole-plot factor (A), one subplot factor (B), and one interaction term (AB). Each factor is defined at two levels, a high setting (1) and a low setting (-1).

   The design matrix, X, for Design 1 example is.

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

   For matrix X, Column 1 represents the intercept; Column 2 represents the setting for Factor A (1 or -1); Column 3 represents the setting for Factor B (1 or -1); Column 4 represents the setting for Interaction AB (1 or -1). The augmented design point, X_1, is defined by row 1 of X, X_2 is defined by row 2 of X, X_3 is defined by row 3 of X, X_4 is defined by row 4 of X, etc.

2. Define C:

   The model for Design 1, includes only four coefficients; the intercept coefficient, Factor A coefficient, Factor B coefficient, and Interaction AB coefficient.

$$C = \begin{pmatrix} 50 \\ 10 \\ 5 \\ 2 \end{pmatrix}.$$

For matrix C, Row 1 represents the coefficient for the Intercept; Row 2 represents the coefficient for Factor A; Row 3 represents the coefficient for Factor B; Row 4 represents the coefficient for Interaction AB;

3. Define E:

   E is a matrix that contains abr (defined previously) elements of $E_{\mathbf{k}}$. For this example there are 12 elements in matrix E. Each $E_{\mathbf{k}}$ is obtained as indicated in Equation 3.1.

$$E = \begin{pmatrix} E_{\mathbf{1}} \\ E_{\mathbf{2}} \\ E_{\mathbf{3}} \\ E_{\mathbf{4}} \\ E_{\mathbf{5}} \\ E_{\mathbf{6}} \\ E_{\mathbf{7}} \\ E_{\mathbf{8}} \\ E_{\mathbf{9}} \\ E_{\mathbf{10}} \\ E_{\mathbf{11}} \\ E_{\mathbf{12}} \end{pmatrix} = \begin{pmatrix} \varepsilon_{\mathbf{1}} + \varepsilon_{\mathbf{11}} \\ \varepsilon_{\mathbf{1}} + \varepsilon_{\mathbf{12}} \\ \varepsilon_{\mathbf{2}} + \varepsilon_{\mathbf{21}} \\ \varepsilon_{\mathbf{2}} + \varepsilon_{\mathbf{22}} \\ \varepsilon_{\mathbf{3}} + \varepsilon_{\mathbf{31}} \\ \varepsilon_{\mathbf{3}} + \varepsilon_{\mathbf{32}} \\ \varepsilon_{\mathbf{4}} + \varepsilon_{\mathbf{41}} \\ \varepsilon_{\mathbf{4}} + \varepsilon_{\mathbf{42}} \\ \varepsilon_{\mathbf{5}} + \varepsilon_{\mathbf{51}} \\ \varepsilon_{\mathbf{5}} + \varepsilon_{\mathbf{52}} \\ \varepsilon_{\mathbf{6}} + \varepsilon_{\mathbf{61}} \\ \varepsilon_{\mathbf{6}} + \varepsilon_{\mathbf{62}} \end{pmatrix}.$$

   For this example, $\varepsilon_{\mathbf{i}} \sim Norm(0, 6)$ and $\varepsilon_{\mathbf{ij}} \sim Norm(0, 10)$ (not a distribution set in Table 3.2). Eighteen random draws are obtained from the two distributions, six ($ar = 2*3$) for $\varepsilon_{\mathbf{i}}$, and 12 ($abr = 2*2*3$) for $\varepsilon_{\mathbf{ij}}$. Therefore, for the example,

$$E = \begin{pmatrix} -2.5954 - 4.3256 \\ -2.5954 - 16.6558 \\ -9.9935 + 1.2533 \\ -9.9935 + 2.8768 \\ 0.7520 - 11.4647 \\ 0.7520 + 11.9092 \\ 1.7261 + 11.8916 \\ 1.7261 - 0.3763 \\ -6.8788 + 3.2729 \\ -6.8788 + 1.7464 \\ 7.1455 - 1.8671 \\ 7.1455 + 7.2579 \end{pmatrix} = \begin{pmatrix} -6.9210 \\ -19.2512 \\ -8.7402 \\ -7.1167 \\ -10.7127 \\ 12.6611 \\ 13.6177 \\ 1.3497 \\ -3.6059 \\ -5.1324 \\ 5.2784 \\ 14.4034 \end{pmatrix}.$$

4. Evaluate Y:

Simple matrix computations are used to generate each Y, as follows:

$$Y = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \times \begin{pmatrix} 50 \\ 10 \\ 5 \\ 2 \end{pmatrix} + \begin{pmatrix} -6.9210 \\ -19.2512 \\ -8.7402 \\ -7.1167 \\ -10.7127 \\ 12.6611 \\ 13.6177 \\ 1.3497 \\ -3.6059 \\ -5.1324 \\ 5.2784 \\ 14.4034 \end{pmatrix} = \begin{pmatrix} 60.0790 \\ 33.7488 \\ 34.2598 \\ 29.8833 \\ 56.2873 \\ 65.6611 \\ 56.6177 \\ 38.3497 \\ 63.3941 \\ 47.8676 \\ 48.2784 \\ 51.4034 \end{pmatrix}.$$

Matrix Y represents the observations from an experiment and serve as the pseudo-experiment results. Throughout each simulation, for each particular bootstrap method, the Matrix Y remains the same. The use of the same Matrix Y for each simulation is synonymous to using the same Common Random Number (CRN) stream. The differences found in comparing the simulations is thus due to the particular bootstrap method used in the analysis rather than due to a difference in Matrix Y.

3.5 Split-plot Analysis

3.5.1 Expected Value Simulation. The expected value (EV) simulation provides the traditional and expected value results for the split-plot analysis. EV verifies the coded split-plot analysis algorithms within MatLab, and validates the expected value theory for split-plot analysis, particularly in regards to whole-plot and subplot error components. The split-plot algorithms used in EV are the same algorithms used in all other simulations.

Within the simulation, standard split-plot analysis is performed. The results are dependent upon both the number of replications performed in the pseudo-experiment and the distribution with which the two errors were generated to form Matrix Y. In addition, as the number of replications increase the whole-plot and subplot error estimates converge to the theoretical expected values; $\sigma_0^2 + b\sigma_1^2$ and σ_0^2, respectively.

The error estimates generated in EV are compared to various bootstrap method estimates. The comparison indicates how well each method captures the known true error components and how much, if at all, bootstrapping improves the error component estimates.

Experiments are simulated to represent experiments providing 2 to 20 replications. This range is used to gain insight into whether bootstrapping use improves with an increase in actual observation set size. In addition, expected value of the simulation is confirmed with the conduct of a larger replicated experiment.

To explain EV, the analysis of Matrix Y is detailed. In the analysis, a represents the number of whole-plots in a single replication of Design 1 ($a = 2$); b represents the number of subplots within each whole plot ($b = 2$); r represents the number of replications observed ($r = 3$). Note that when the design analyzed is of differing size than that in Design 1, a and b are determined by the following:

$$a = \#\text{ whole-plot factors } \times \#\text{ whole-plot levels}$$

$$b = \#\text{ subplot factors } \times \#\text{ subplot levels}$$

Split-Plot analysis on Matrix Y determines estimates for the whole-plot error and subplot error. The results of the analysis are the following:

The sums of squares for the whole-plot terms are as follows:

$SS_{replicate} = \sum \frac{Y_{i..}^2}{ab} - \frac{Y_{...}^2}{abr}$; for $i = 1, 2, 3$

$= 526.3576$

$SS_{FactorA} = \sum \frac{Y_{.j.}^2}{br} - \frac{Y_{...}^2}{abr}$; for $j = 1, 2$

$= 388.1208$

$SS_{WPerror} = SS_{WP} - SS_{FactorA} - SS_{replicate}$

$= 47.6907$

From the $SS_{WPerror}$ the estimate for whole plot error is:

$MS_{WPerror} = \frac{SS_{WPerror}}{(r-1)(a-1)}$

$= 23.8454$

The sums of squares for the subplot terms are:

$SS_{FactorB} = \sum \frac{Y_{..k}^2}{ar} - \frac{Y_{...}^2}{abr}$; for $k = 1, 2$

$= 225.3542$

$SS_{AB} = \sum \frac{Y_{.jk}^2}{r} - \frac{Y_{...}^2}{abr} - SS_{FactorA} - SS_{FactorB}$; for $j, k = 1, 2$

$= 14.004$

$SS_{SPerror} = SS_{Total} - SS_{replicate} - SS_{FactorA} - SS_{WPerror} - SS_{FactorB} - SS_{AB}$

$= 453.0722$

From the $SS_{SPerror}$ the estimate for the Subplot error is:

$MS_{SPerror} = \frac{SS_{SPerror}}{a(r-1)(b-1)}$

$= 113.2680$

The whole-plot error estimate is 23.8454 and the subplot error estimate is 113.2680. The true whole-plot error is 172 while the true subplot error estimate is 100. This is a problem with small samples. However, as the number of replications increase, both the whole-plot error and subplot error estimate converge to the true values. At 10,000 replications, the whole plot error estimate is 172.8368 and the subplot error estimate is 99.9797. This example illustrates how a small number of replications may not be enough. The next question is whether the sample size is sufficient for a bootstrap approach to improve the error estimate. Thus, these error estimates are compared to various re-sampling methods for each combination of design, distribution set, and experimental replication.

3.6 Bootstrap Methods

Five separate re-sampling methods are examined, three based on residual re-sampling, two based on re-sampling the pseudo-experiment observations. The residual methods, RM1, RM2, RM3, vary how residuals are re-sampled with respect to whole-plot or subplot error structure. The observational methods, OM1 and OM2, vary how multiple replication pseudo-experiments are re-sampled.

3.6.1 Bootstrap Simulation–Residual Method 1. The Residual Method 1 (RM1) simulation employs the CBRM methodology. It begins with the initial observations (Matrix Y) and fits a linear regression model using Least Squares methodology,

$$C^* = (X'X)^{-1}X'Y. \tag{3.4}$$

Using the Design 1 example,

$$X'X = \begin{pmatrix} 12 & 0 & 0 & 0 \\ 0 & 12 & 0 & 0 \\ 0 & 0 & 12 & 0 \\ 0 & 0 & 0 & 12 \end{pmatrix}; \ (X'X)^{-1} = \begin{pmatrix} \frac{1}{12} & 0 & 0 & 0 \\ 0 & \frac{1}{12} & 0 & 0 \\ 0 & 0 & \frac{1}{12} & 0 \\ 0 & 0 & 0 & \frac{1}{12} \end{pmatrix};$$

$$X'Y = \begin{pmatrix} 591.2892 \\ 80.6231 \\ 53.6019 \\ 15.1707 \end{pmatrix}; \ C^* = \begin{pmatrix} 49.2741 \\ 6.7186 \\ 4.4668 \\ 1.2642 \end{pmatrix}.$$

The observations are formed by matrix multiplication of the augmented design points and the newly found regression coefficients,

$$Y_{fit} = X \times C^*. \tag{3.5}$$

$$Y_{fit} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \times \begin{pmatrix} 48.8192 \\ 5.6871 \\ 4.3335 \\ 1.0803 \end{pmatrix} = \begin{pmatrix} 59.9201 \\ 49.0925 \\ 46.3853 \\ 39.8788 \\ 59.9201 \\ 49.0925 \\ 46.3853 \\ 39.8788 \\ 59.9201 \\ 49.0925 \\ 46.3853 \\ 39.8788 \end{pmatrix}.$$

The residuals are obtained by subtracting the fitted observations from the initial observations,

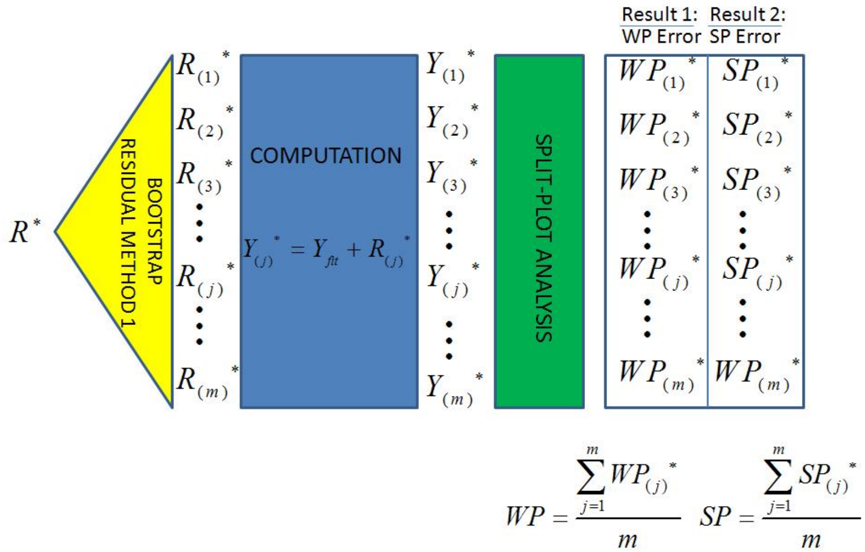


Figure 3.3: Bootstrap Residual Method 1 Schematic.

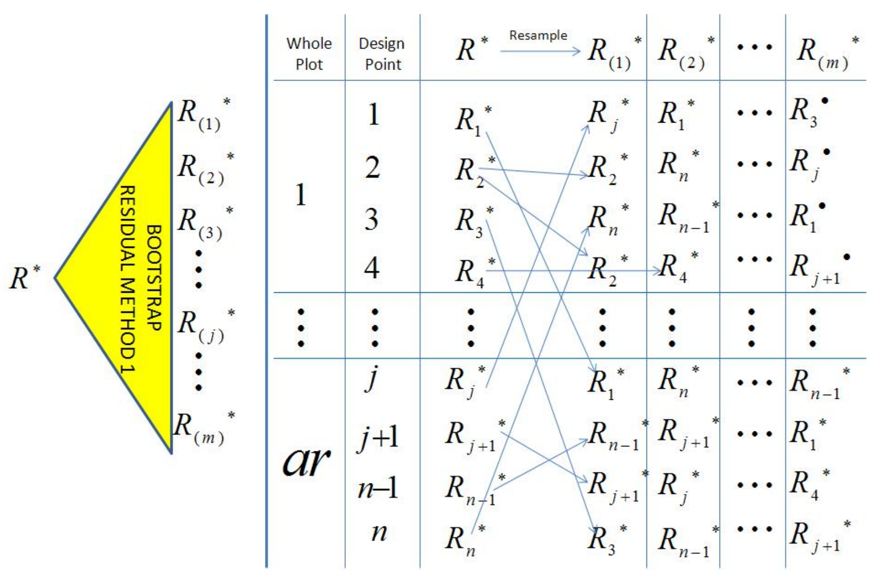


Figure 3.4: Residual Method 1 Bootstrap Methodology Details.

$$R^* = Y - Y_{fit}. \tag{3.6}$$

The schematic for the methodology employed in RM1 is shown in Figure 3.3. The details of the specific bootstrap technique used are represented in Figure 3.4.

$$R^* = \begin{pmatrix} 60.0790 \\ 33.7488 \\ 34.2598 \\ 29.8833 \\ 56.2873 \\ 65.6611 \\ 56.6177 \\ 38.3497 \\ 63.3941 \\ 47.8676 \\ 48.2784 \\ 51.4034 \end{pmatrix} - \begin{pmatrix} 59.9201 \\ 49.0925 \\ 46.3853 \\ 39.8788 \\ 59.9201 \\ 49.0925 \\ 46.3853 \\ 39.8788 \\ 59.9201 \\ 49.0925 \\ 46.3853 \\ 39.8788 \end{pmatrix} = \begin{pmatrix} 0.1589 \\ -15.3437 \\ -12.1255 \\ -9.9955 \\ -3.6328 \\ 16.5687 \\ 10.2324 \\ -1.5291 \\ 3.4740 \\ -1.2249 \\ 1.8931 \\ 11.5246 \end{pmatrix}.$$

The residuals (R^*) are then re-sampled with replacement to generate a bootstrapped sample of residuals, $R^*_{(j)}$; 1000 such samples are generated. In RM1, each residual has an equal chance being sampled, but it may not occur in each bootstrap sample (e.g., R^*_4 occurred in bootstrap sample $R^*_{(2)}$, but not $R^*_{(1)}$). A specific residual can also be repeated within a bootstrap sample, such as R^*_2, sampling with replacement is used. An important concept in RM1 is that the re-sampling does not factor in whether the residual is whole plot or subplot. RM1 method omits the dependent structure of the observations [3]. Although R^*_1 is the residual for design point 1 within whole plot 1, R^*_1 can occur in the bootstrap sample in any whole plot. Figure 3.4 indicates this for bootstrap sample $R^*_{(1)}$ when R^*_1 becomes the residual for design point j which is in whole-plot ar. Also, every bootstrap sample includes the same number

Table 3.3: $R^*_{(j)}$ example for RM1

$R^*_{(1)}$	$R^*_{(2)}$	$R^*_{(3)}$	...	$R^*_{(1000)}$
-1.2249	11.5246	3.4740	...	3.4740
1.8931	16.5687	-1.2249	...	-12.1255
-15.3437	-1.2249	3.4740	...	16.5687
1.8931	-15.3437	-3.6328	...	0.1589
-1.5291	16.5687	-1.5291	...	16.5687
-15.3437	1.8931	-12.1255	...	-9.9955
-9.9955	-1.2249	3.4740	...	3.4740
10.2324	11.5246	0.1589	...	3.4740
11.5246	-1.5291	-9.9955	...	-3.6328
11.5246	0.1589	0.1589	...	10.2324
-15.3437	1.8931	-15.3437	...	3.4740
11.5246	11.5246	-1.2249	...	-1.5291

of observations as in the initial observations, Y. If r replications are in Matrix Y, then r replications are produced for $R^*_{(j)}$.

Since the methodology produces values for $R^*_{(j)}$ such that j = 1,2,...,1000, only $R^*_{(j)}$ values for j = 1, 2, 3, and 1000 are provided in Table 3.3 .

The new observations are generated by:

$$Y^*_{(j)} = Y_{fit} + R^*_{(j)} \tag{3.7}$$

$Y^*_{(j)}$ values for j = 1, 2, 3, and 1000 are provided in Table 3.4 .

Split-plot analysis is performed on the new sample of observations, $Y^*_{(j)}$. The whole plot and subplot error estimates are recorded. 1,000 bootstrap samples are generated to obtain 1,000 estimates for the whole plot and subplot errors. The whole plot and subplot error estimates are then aggregated to obtain a point estimate for the two errors using WP and SP equations represented in Figure 3.3. The aggregated estimates are then compared to the true values and to values obtained via EV.

The error estimates for j = 1, 2, 3, and 1000 are provided in Table 3.5 .

Table 3.4: $Y^*_{(j)}$ example for RM1

$Y^*_{(1)}$	$Y^*_{(2)}$	$Y^*_{(3)}$	...	$Y^*_{(1000)}$
58.6952	71.4447	63.3941	...	63.3941
50.9856	65.6611	47.8676	...	36.9670
31.0416	45.1604	49.8593	...	62.9540
41.7719	24.5351	36.2460	...	40.0376
58.3910	76.4888	58.3910	...	76.4888
33.7488	50.9856	36.9670	...	39.0970
36.3898	45.1604	49.8593	...	49.8593
50.1112	51.4034	40.0376	...	43.3528
71.4447	58.3910	49.9246	...	56.2873
60.6171	49.2513	49.2513	...	59.3249
31.0416	48.2784	31.0416	...	49.8593
51.4034	51.4034	38.6539	...	38.3497

Table 3.5: Bootstrap Error Estimates for RM1

$Y^*_{(j)}$	Whole Plot Error	Subplot Error
$j = 1$	127.9929	26.3788
$j = 2$	224.7176	81.8165
$j = 3$	40.9805	60.6184
$j = 1000$	64.6885	126.9732

3.6.2 Bootstrap Simulation–Residual Method 2. Residual Method 2 (RM2) simulation employs the CBRM methodology and therefore begins with the initial observations (Matrix Y) and fits a linear regression model using equation 3.4 to yield

$$C^* = \begin{pmatrix} 49.2741 \\ 6.7186 \\ 4.4668 \\ 1.2642 \end{pmatrix}.$$

The newly fitted observations are evaluated by equation 3.5

$$Y_{fit} = \begin{pmatrix} 59.9201 \\ 49.0925 \\ 46.3853 \\ 39.8788 \\ 59.9201 \\ 49.0925 \\ 46.3853 \\ 39.8788 \\ 59.9201 \\ 49.0925 \\ 46.3853 \\ 39.8788 \end{pmatrix}.$$

The residuals are obtained by equation 3.6

The schematic for the methodology employed in RM2 is shown in Figure 3.5. The details of the specific bootstrap technique used is represented in Figure 3.6.

Since the methodology produces values for $R^{\bullet}_{(j)}$ such that $j = 1,2,\ldots,1000$, only $R^{\bullet}_{(j)}$ values for $j = 1, 2, 3,$ and 1000 are provided in Table 3.8 .

The residuals in each specific whole plot are resampled with replacement to generate a bootstrapped sample of residuals, $R^{\bullet}_{(j)}$. Each residual within a whole plot has an equal chance of occuring for an observation within that whole plot. Therefore,

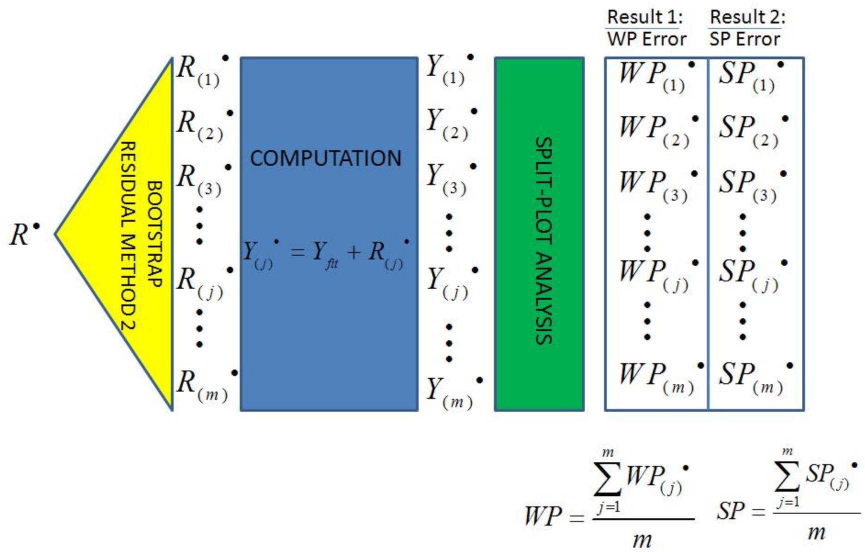


Figure 3.5: Bootstrap Residual Method 2 Schematic.

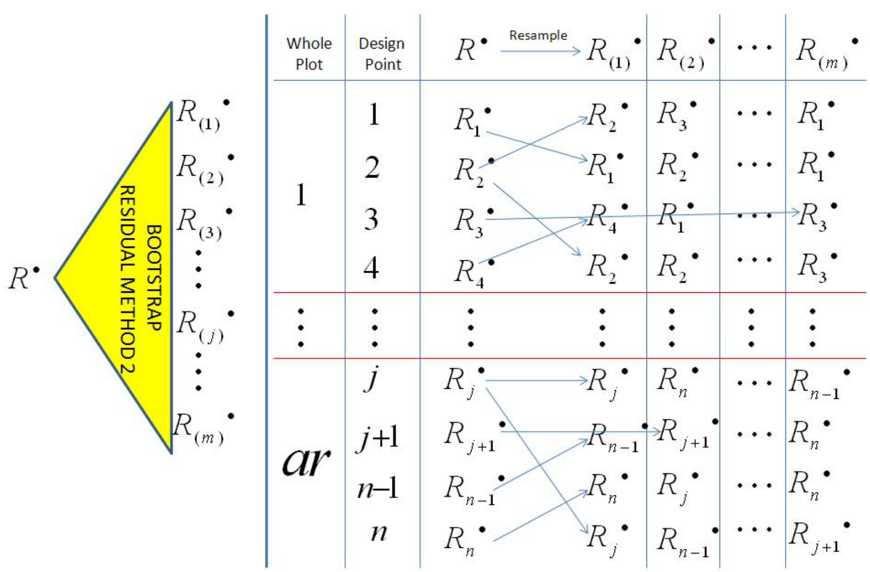


Figure 3.6: Residual Method 2 Bootstrap Methodology Details.

Table 3.6: $R^{\bullet}_{(j)}$ example for RM2

$R^{\bullet}_{(1)}$	$R^{\bullet}_{(2)}$	$R^{\bullet}_{(3)}$	...	$R^{\bullet}_{(1000)}$
-1.2249	-15.3437	10.2324	...	-12.1255
-1.5291	-1.2249	-12.1255	...	0.1589
-9.9955	-9.9955	-1.2249	...	-3.6328
1.8931	0.1589	10.2324	...	-1.2249
-1.5291	-12.1255	-3.6328	...	3.4740
-12.1255	3.4740	-9.9955	...	11.5246
3.4740	1.8931	-15.3437	...	0.1589
-1.5291	-3.6328	-1.2249	...	1.8931
1.8931	10.2324	11.5246	...	-1.5291
10.2324	-9.9955	3.4740	...	10.2324
11.5246	0.1589	16.5687	...	-1.5291
-15.3437	3.4740	-9.9955	...	10.2324

it does matter what whole plot the residual comes from. This method attempts to address the dependence among observations within a whole-plot [3]. If $R^{\bullet}_1$ is a residual for a design point in whole-plot 1, $R^{\bullet}_1$ can only occur as a residual in a bootstrap sample for a design point within whole-plot 1. A residual can repeat within a bootstrap sample as represented by $R^{\bullet}_j$ in Figure 3.6. In addition, every bootstrap sample will include the same number of observations as in the initial observations.

The new observations are then generated by equation 3.7.

Split-plot analysis is performed on the new sample of observations, $Y^{\bullet}_{(j)}$. The whole-plot and subplot error estimates are recorded. 1,000 bootstrap samples are generated to obtain 1,000 estimates for the whole-plot and subplot errors. The whole-plot and subplot error estimates are then aggregated to obtain a point estimate for the two errors using WP and SP equations in Figure 3.5. The aggregated estimates are then compared to the true values and to values obtained via EV.

3.6.3 Bootstrap Simulation–Residual Method 3. Residual Method 3 (RM3) simulation employs the CBRM methodology, begins with the initial observations (Matrix Y), and fits a linear regression model using equation 3.4 to yield

Table 3.7: $Y^{\bullet}_{(j)}$ example for RM2

$Y^{\bullet}_{(1)}$	$Y^{\bullet}_{(2)}$	$Y^{\bullet}_{(3)}$	...	$Y^{\bullet}_{(1000)}$
57.0242	42.9054	68.4815	...	46.1236
50.373	50.6772	39.7766	...	52.061
29.4858	29.4858	38.2564	...	35.8485
45.9289	44.1947	54.2682	...	42.8109
56.72	46.1236	54.6163	...	61.7231
39.7766	55.3761	41.9066	...	63.4267
42.9553	41.3744	24.1376	...	39.6402
42.5067	40.403	42.8109	...	45.9289
60.1422	68.4815	69.7737	...	56.72
62.1345	41.9066	55.3761	...	62.1345
51.0059	39.6402	56.05	...	37.9522
28.6921	47.5098	34.0403	...	54.2682
[1ex] height				

$$C^* = \begin{pmatrix} 49.2741 \\ 6.7186 \\ 4.4668 \\ 1.2642 \end{pmatrix}.$$

The fitted observations are evaluated by equation 3.5

$$Y_{fit} = \begin{pmatrix} 59.9201 \\ 49.0925 \\ 46.3853 \\ 39.8788 \\ 59.9201 \\ 49.0925 \\ 46.3853 \\ 39.8788 \\ 59.9201 \\ 49.0925 \\ 46.3853 \\ 39.8788 \end{pmatrix}.$$

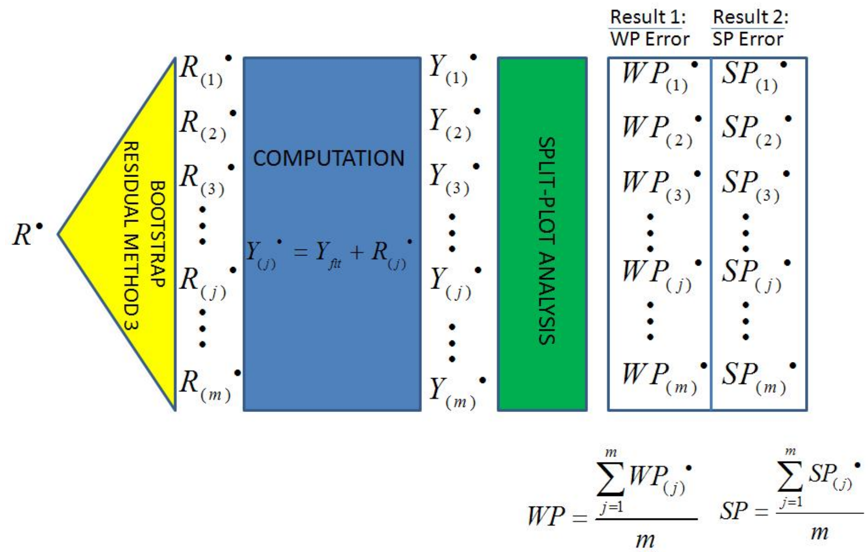


Figure 3.7: Bootstrap Residual Method 3 Schematic.

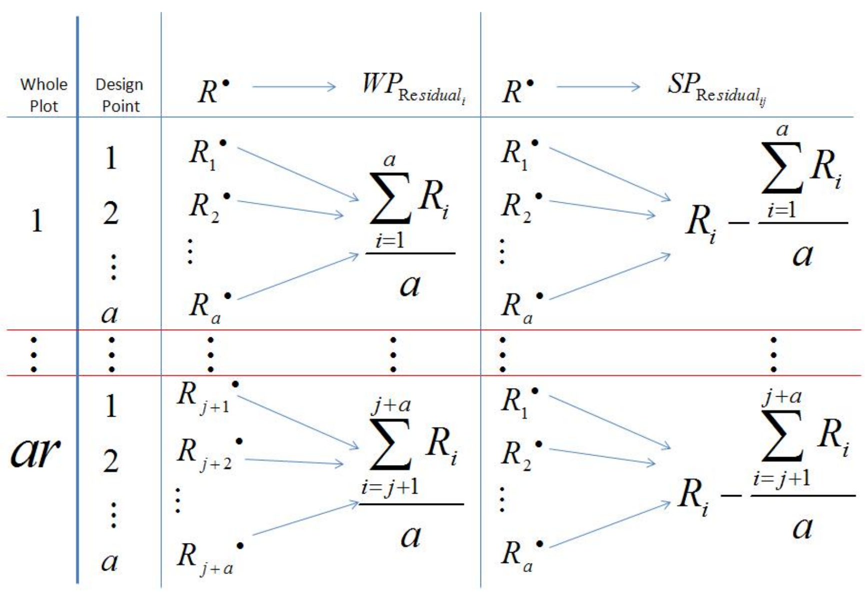


Figure 3.8: Residual Method 3 Bootstrap Methodology Details.

Table 3.8: $R^{\bullet}_{(j)}$ example for RM3

$R^{\bullet}_{(1)}$	$R^{\bullet}_{(2)}$	$R^{\bullet}_{(3)}$	...	$R^{\bullet}_{(1000)}$
-6.9259	3.8477	-14.3705	...	6.3537
-4.9632	7.2875	-18.8393	...	7.6266
11.8945	1.3872	-8.7422	...	0.1817
10.3498	-0.4291	-5.6529	...	2.0493
0.8439	13.2740	6.3537	...	-18.8393
2.6601	11.6504	5.3923	...	-14.3705
6.7718	3.7490	4.9545	...	2.8152
5.1482	3.3309	2.8152	...	4.7104
2.4761	-15.0603	0.8439	...	-7.8873
5.2261	-14.3705	2.3210	...	-5.9921
-15.6711	-8.4030	13.2740	...	12.3126
-17.8104	-6.2637	11.6504	...	11.0396

The residuals are obtained by equation 3.6. However, instead of using bootstrap procedures as in RM2, RM3 will estimate the whole-plot error and subplot error for each individual observation and then bootstraps on whole-plot error as well as subplot error. The whole-plot error estimation is obtained by averaging the residuals within whole-plot. This average then becomes the whole-plot error estimate for all observations within this whole-plot. The subplot estimates are found by subtraction.

The schematic for the methodology employed in RM3 is shown in Figure 3.5. The details of the specific bootstrap technique used is represented in Figure 3.6.

Since the methodology produces values for $R^{\bullet}_{(j)}$ such that $j = 1,2,\ldots,1000$, only $R^{\bullet}_{(j)}$ values for $j = 1$, 2, 3, and 1000 are provided in Table 3.8 .

The whole-plot error residuals are re-sampled with replacement to generate bootstrapped samples of whole-plot residuals. The structure of the whole-plot error residuals does matter therefore this method attempts to address the correlation among observations within a whole-plot by maintaining the same whole-plot error residual for each observation within a whole-plot [3]. The subplot error residuals are also re-sampled with replacement to generate bootstrap samples of subplot residuals. Once the whole-plot and subplot residuals are generated, the whole-plot and subplot

Table 3.9: $Y^{\bullet}_{(j)}$ example for RM3

$Y^{\bullet}_{(1)}$	$Y^{\bullet}_{(2)}$	$Y^{\bullet}_{(3)}$	...	$Y^{\bullet}_{(1000)}$
62.9134	64.1189	56.2239	...	76.2497
47.5702	49.5329	40.2860	...	60.3395
45.3899	27.5144	49.2675	...	46.6629
42.2804	22.0730	43.1363	...	41.3190
56.4680	67.8799	73.0815	...	62.6417
41.2198	55.7049	59.7287	...	47.5702
48.3061	33.4819	44.9632	...	50.5405
45.6147	29.5851	38.8320	...	44.7599
47.3323	77.5503	47.3323	...	46.3985
30.8787	59.7287	31.3944	...	34.3185
56.1879	46.2447	49.1689	...	34.1717
52.8067	43.5534	43.5534	...	31.9960

residuals are added to produce the new residual for the bootstrap samples. The new observations are then generated by equation 3.7.

Split-plot analysis is performed on the new sample of observations, $Y^{\bullet}_{(j)}$. The whole-plot and subplot error estimates are recorded. 1,000 bootstrap samples are generated to obtain 1,000 estimates for the whole-plot and subplot errors. The whole-plot and subplot error estimates are then aggregated to obtain a point estimate for the two errors using WP and SP equations in Figure 3.5. The aggregated estimates are then compared to the true values and to values obtained via EV.

3.6.4 Bootstrap Simulation–Obervations Method 1. The Observations Method 1 (OM1) simulation begins with the initial set of observations (Matrix Y). The observations are sampled with replacement across replications to generate a bootstrapped sample of observations, $Y^{\odot}_{(j)}$. The schematic for the methodology employed in OM1 is shown in Figure 3.9. The details of the specific bootstrap technique used is represented in Figure 3.10.

Each replicated observation associated with a design point has an equal chance of occurring. In addition, every bootstrap sample will include the same number of

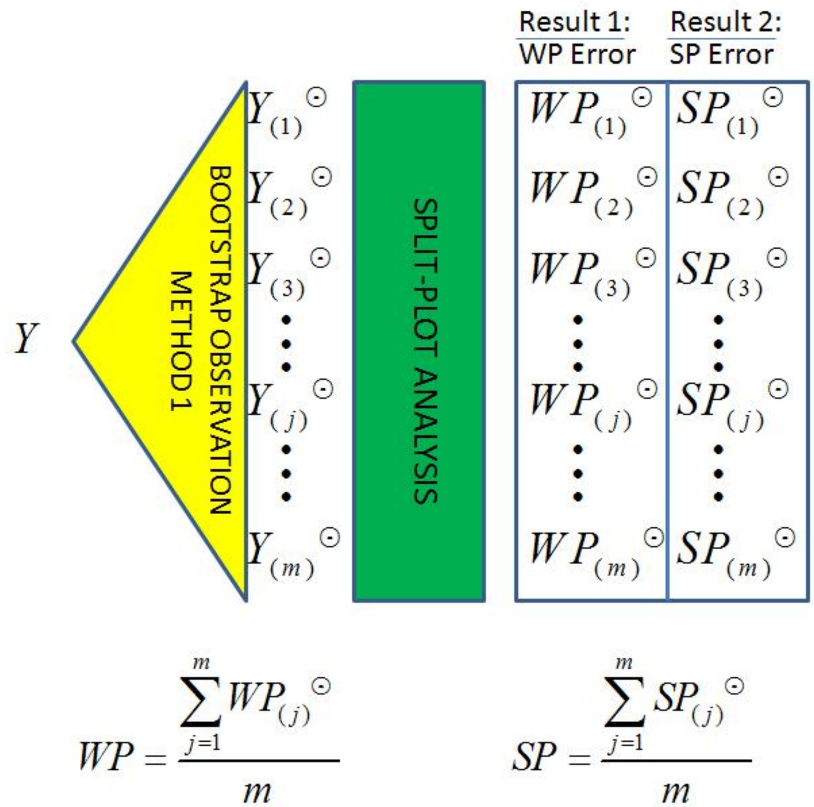


Figure 3.9: Bootstrap Observation Method 1 Schematic.

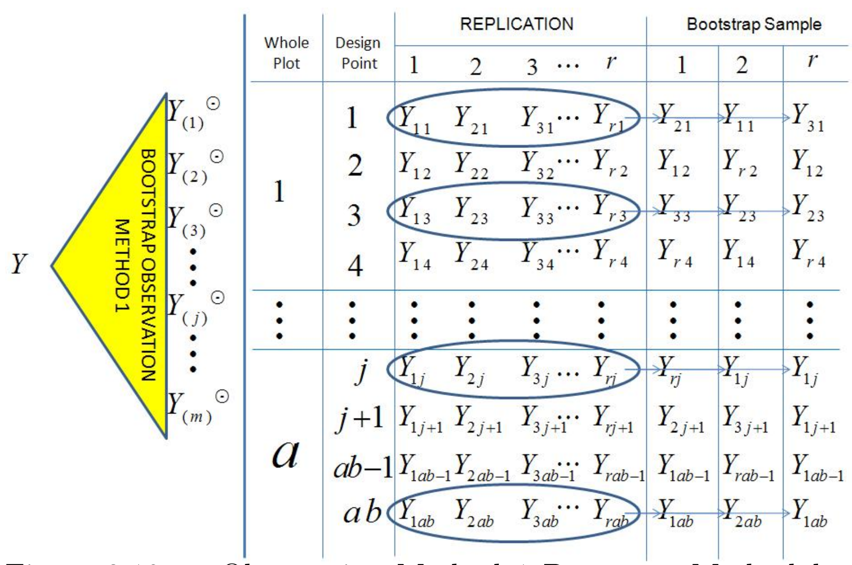


Figure 3.10: Observation Method 1 Bootstrap Methodology Details.

observations as in the initial observations, Matrix Y. If r replications are in Matrix Y, then r replications are provided for bootstrap sample $Y^{\odot}_{(j)}$

Split-plot analysis is performed on the new sample of observations, $Y^{\odot}_{(j)}$. The whole plot and subplot error estimates are recorded. 1,000 bootstrap samples are generated to obtain 1,000 estimates for the whole plot and subplot errors. The whole plot and subplot error estimates are then aggregated to obtain a point estimate for each error estimate. The aggregated estimates are then compared to the true values and to values obtained via EV.

3.6.5 Bootstrap Simulation–Obervations Method 2. The Observations Method 2 (OM2) simulation begins with the initial observations (Matrix Y) The observations of a specific design point are sampled with replacement across replication to generate a bootstrapped sample of observations, $Y^{\circ}_{(j)}$. The bootstrap sample formed has 250 replications rather than the r replications represented in Matrix Y. The schematic for the methodology employed in OM2 is shown in Figure 3.11. The details of the specific bootstrap technique used is represented in Figure 3.12. Split-plot analysis is performed on the new sample of observations, $Y^{\circ}_{(j)}$. The whole plot and subplot error estimates are recorded and then compared to true values and to values obtained via EV.

3.7 Comparison Criteria

Comparison methods are used to assess how well, if at all, the bootstrap methods improve error estimation in split-plot analysis. Three methods are used in this research. The results chapter uses the first method primarily with details on the sign test and paired-t test results provided in Appendix A.

3.7.1 Direct Comparison. The benefit of a Monte Carlo study is that the true error components are known. Thus, the primary measure of comparison employed is how well EV and each bootstrap method estimates the true error structure

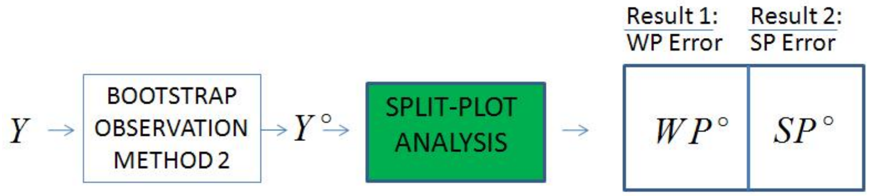


Figure 3.11: Bootstrap Observation Method 2 Schematic.

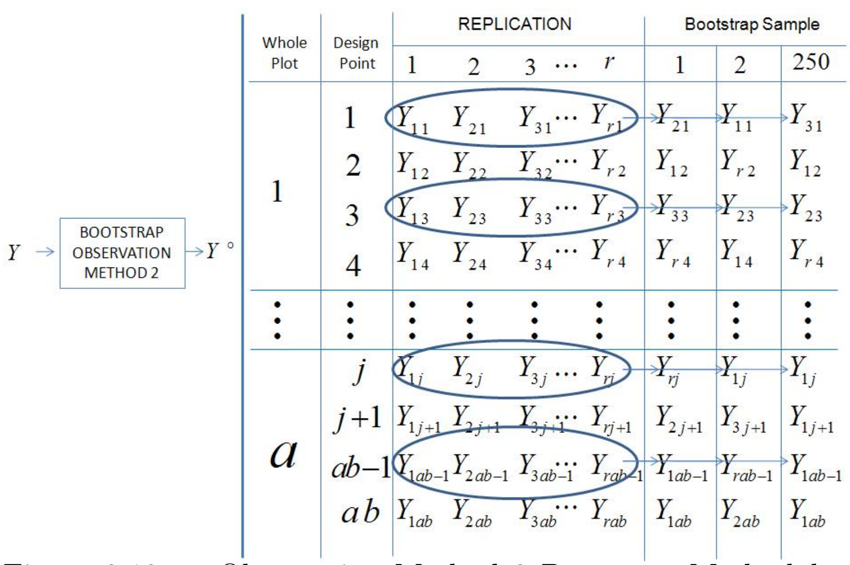


Figure 3.12: Observation Method 2 Bootstrap Methodology Details.

components. This comparison involves all 5 re-sampling methods, the EV method, across 9 split-plot designs, 13 distributional sets, and 2 to 20 replications per pseudo-experiment. In subsequent analyses, the data presented are restricted to the 9 split-plot designs, across 3 distributional sets, for 2, 5, 10, and 20 replications per pseudo-experiment.

Define $m_j, j = 1, \ldots, 5$ as methods RM1, RM2, RM3, OM1 and OM2, respectively. Let T denote the true error component and EV the associated EV estimate. Then, let WP_{m_j} and SP_{m_j} represent the whole-plot and subplot, respectively, error estimate obtained via method j. Let WP_T, WP_T, WP_{EV}, and SP_{EV} represent the corresponding true error and EV-estimated error values. Assume each Monte Carlo experiment is replicated K times. Then,

$$d_{1k} = WP_{m_j} - WP_T, \qquad k = 1, \ldots, K; j = 1, \ldots, 5. \tag{3.8}$$

$$d_{2k} = SP_{m_j} - SP_T, \qquad k = 1, \ldots, K; j = 1, \ldots, 5. \tag{3.9}$$

$$d_{3k} = WP_{EV} - WP_T, \qquad k = 1, \ldots, K; j = 1, \ldots, 5. \tag{3.10}$$

$$d_{4k} = SP_{EV} - SP_T, \qquad k = 1, \ldots, K; j = 1, \ldots, 5. \tag{3.11}$$

Calculate means, $\bar{d}_1, \bar{d}_2, \bar{d}_3,$ and $\bar{d}_4$ from the $d_{1k}, d_{2k}, d_{3k},$ and d_{4k} data, respectively, and produce confidence intervals for each mean reported by split-plot design and distribution set, for each level of pseudo-experiment replication.

These data yield insight into how accurate each of EV and the re-sampling methods estimate the true error structure as a function of design size, error structure, and replication level.

3.7.2 Sign test. The sign test is a non-parametric test based on the binomial distribution and used to determine whether two samples (X and Y) are represented by the same underlying distribution. If the two samples are from the same distribution, then $x_i \in X$ and $y_i \in Y$ are equally likely to be larger than the other. Therefore, to

use the sign test, the number of times x_i is larger than y_i is counted and denoted as w (ties are ignored). The probability that at least w wins occur ($p = P(W \geq w)$) by chance alone is then represented by the binomial distribution, $W = bin(n, 0.5)$. In this case, n is the number of non-equal valued comparisons available.

The use of the sign test in this research is adapted as such:

1. Whole plot (wp_i) and subplot (sp_i) error component estimates are found for two simulated methods; $i = 1$ represents method 1, $i = 2$ represents method 2; with method one generally a re-sampling method and method 2 the EV method.

2. Accuracy of error estimates based on theory is determined, $A_{wp_i} = |E(wp) - wp_i|$ and $A_{sp_i} = |E(sp) - sp_i|$, respectively; where

$$E(wp) = \sigma_0^2 + b\sigma_1^2 \tag{3.12}$$

and

$$E(sp) = \sigma_0^2 \tag{3.13}$$

3. Minimum value is determined

$$min_{wp} = min(A_{wp_1}, A_{wp_2}) \tag{3.14}$$

$$min_{sp} = min(A_{sp_1}, A_{sp_2}) \tag{3.15}$$

4. Determine count, w, based on following: If $min_w p = A_{wp_1}$ increase w by 1

5. Determine p-value of the sign test

$$p - value = P(W \geq w) \tag{3.16}$$

Note: If the p-value is less than α, method 1 is more accurate method. However, if the p-value is greater than $1 - \alpha$, method 2 is more accurate method.

These data results are provided in Appendix A but summarized in the Results chapter.

3.7.3 Paired-t test. A paired-t test is used to formulate a confidence interval that can help determine whether two samples (X and Y) are represented by the same underlying distribution. The difference between the samples is calculated for each pair, $Z = Y - X$. From the differences, the mean ($\bar{Z}$) and standard deviation (σ_Z) are calculated. A confidence interval is formed based on the mean and standard deviation of the differences. If the interval contains 0 then there is not sufficient evidence to conclude the two samples are from different underlying distributions. An assumption with this test is that the differences between the two samples, $Z = Y - X$ are normally distributed.

The use of the paired-t test in this research is adapted as such:

1. Whole plot (wp_i) and subplot (sp_i) error component estimates are found for two simulated methods; $i = 1$ represents method 1, $i = 2$ represents method 2;
2. Accuracy of error estimates based on theory is determined, $A_{wp_i} = \mid E(wp) - wp_i \mid$ and $A_{sp_i} = \mid E(sp) - sp_i \mid$, respectively; where

$$E(wp) = \sigma_0^2 + b\sigma_1^2 \tag{3.17}$$

and

$$E(sp) = \sigma_0^2 \tag{3.18}$$

3. Differences formulated

$$Z_{wp} = A_{wp_1} - A_{wp_2} \tag{3.19}$$

$$Z_{sp} = A_{sp_1} - A_{sp_2} \tag{3.20}$$

4. Mean and standard deviation calculated
5. Confidence interval formed

These results are provided in Appendix A. These results help determine whether any bootstrapping method improves the error component estimate as compared to the EV method. The Results chapter summarizes these data results.

IV. Analysis and Results

This chapter compares the expected value (EV) simulation and the five bootstrap methods to the true model error components. The accuracy and precision of each method is discussed while the sign test and paired-t test results are summarized.

4.1 Simulation Validation and Verification

The EV approach had multiple purposes: to verify the MatLab split-plot analysis algorithms, to validate simulation output, and to generate the standard split-plot analysis for comparison. Verification involved analyzing the data from the experiment in Table 2.2 to confirm that proper results are obtained (Table 2.3). Validation involved performing expected value calculations for the whole-plot and subplot error components for the designs in Table 3.1 and a subset of the distributions in Table 3.2. With the simulation, 10,000 replications are analyzed to provide estimates of the expected value calculations. These values are compared to the true error components for both whole-plot and subplot error, $\sigma_0^2 + b\sigma_1^2$ and σ_0^2, respectively. Validation results are included in Table 4.1.

Table 4.1: Simulation Validation

Split-Plot Design	Distribution Structure	Expected Error - Sim		Expected Error - Theory	
		WP_{error}	SP_{error}	WP_{error}	SP_{error}
1	1	12.0382	3.9992	12	4
2	1	20.0092	4.0405	20	4
3	1	36.3433	4.0159	36	4
4	5	108.6039	100.4642	108	100
5	5	116.4417	100.4317	116	100
6	5	133.5337	100.2209	132	100
7	13	205.3691	3.9891	204	4
8	13	406.4293	3.9917	404	4
9	13	809.1066	3.9925	804	4

The Table 4.1 results indicate that EV is an accurate representation of standard split-plot analysis. The same algorithms are used to estimate the whole-plot and subplot errors for each of the bootstrap methods.

4.2 Direct Comparison

Each re-sampling method and the EV method are compared to the true split-plot error components across 9 split-plot designs, 3 distributional sets, for the 2, 5, 10,

and 20 replication designs. All comparisons are based on $K = 20$. These results are used to determine the merits of each of the 5 bootstrap methods towards improving the split-plot error estimates. Summaries of the results in Appendix A focus on whether bootstrapping helps improve error estimates beyond what the EV method accomplishes. All confidence intervals in the subsequent comparisons method are at an individual $\alpha = 0.05$ level of significance.

4.2.1 EV. The direct comparison confidence intervals (CIs) and mean diference from truth for EV at each design, distributional set and replication level are included in Table 4.2. The results provide estimates of the error components attainable just using the actual test results. The results are an indication of the general robustness, accuracy and precision of split-plot analysis across design, distribution set and replication levels.

In general, EV performs well, clearly, with fewer replications, the error estimates are not as accurate as CI widths are larger (less precise), and fewer design CIs contain the true error components. For the 2 replication designs in the study, only 9 of the 27 subplot error and 22 of the 27 whole-plot error CIs contain the true error component. Designs with 20 replications showed 24 of the 27 subplot and 25 of the 27 whole-plot CIs, contain the true error component. Distribution 5 employs a subplot distribution much larger than the whole-plot distribution, something unlikely to occur in practice. Many of the 2 replication designs that failed to cover the true value were Distribution 5 cases (6 of the 18 subplot and 2 of the 5 whole-plot failures). This empirical evidence indicates that improvements in accuracy and precision of error estimation is warranted,particularly for experiments with fewer replications. Thus, new methods based on re-sampling are investigated in split-plot analysis to determine if they improve the accuracy and precision of the error estimates.

Table 4.2: Expected Value Direct Comparison Confidence Intervals

Design	Distribution	2 replications						5 replications						10 replications						20 replications					
		ΔSP			ΔWP			ΔSP			ΔWP			ΔSP			ΔWP			ΔSP			ΔWP		
		Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean
1	1	-2.8253	-0.9914	-1.90835	-8.786	6.0699	-1.35805	-1.0944	0.2093	-0.44255	-4.5408	-0.3907	-2.46575	-1.0077	0.4965	-0.2556	-2.5968	0.2003	-1.19825	-0.2573	0.5186	0.13065	-2.7448	0.8048	-0.97
1	5	-70.6333	-24.7856	-47.70945	-47.1697	79.1879	16.0091	-27.3595	5.2324	-11.06355	-42.3992	27.2251	-7.58705	-25.192	12.412	-6.39	-32.9925	3.5741	-14.7092	-6.4315	12.9651	3.2668	-12.3109	16.992	2.34055
1	13	-2.8253	-0.9914	-1.90835	-132.2277	137.2898	2.53105	-1.0944	0.2093	-0.44255	-84.0445	-1.2795	-42.662	-1.0077	0.4965	-0.2556	-24.6587	19.8233	-2.4177	-0.2573	0.5186	0.13065	-51.3925	2.9571	-24.2177
2	1	-0.864	0.9051	0.02055	-13.5501	2.3288	-5.61065	-0.7221	0.0772	-0.32245	-7.2977	7.0038	-0.14695	-0.4371	0.1062	-0.16545	-3.2138	4.2458	0.516	-0.2428	0.2189	-0.01195	-2.5145	1.53	-0.49225
2	5	-21.6001	22.6283	0.5141	-79.1626	26.6521	-26.25525	-18.0527	1.9305	-8.0611	-31.9728	37.4154	2.7213	-10.9279	2.6545	-4.1367	-29.657	26.4622	-1.5974	-6.0696	5.4722	-0.2987	-12.2371	19.0578	3.41035
2	13	-0.864	0.9051	0.02055	-229.857	31.9337	-98.96165	-0.7221	0.0772	-0.32245	-137.7516	98.3807	-19.68545	-0.4371	0.1062	-0.16545	-50.2776	134.4446	42.0835	-0.2428	0.2189	-0.01195	-60.5579	37.2719	-11.643
3	1	-0.0532	1.3599	0.65335	-25.491	7.0053	-9.24285	0.0126	0.6978	0.3552	-13.1859	4.2912	-4.44735	-0.1101	0.3908	0.14035	-9.3761	4.4409	-2.4676	-0.0567	0.254	0.09865	-7.3703	1.2382	-3.06605
3	5	-1.3312	33.9966	16.3327	-99.2921	-31.4013	-65.3467	0.3154	17.4461	8.88075	-29.2172	53.647	12.2149	-2.7515	9.7696	3.50905	-30.9609	23.5494	-3.70575	-1.4168	6.3495	2.46635	-6.3641	25.0847	9.3603
3	13	-0.0532	1.3599	0.65335	-531.5593	196.3438	-167.60775	0.0126	0.6978	0.3552	-355.9881	-6.0591	-181.0236	-0.1101	0.3908	0.14035	-163.253	143.1887	-10.0321	-0.0567	0.254	0.09865	-196.123	2.4194	-96.8517
4	1	-0.6953	1.06	0.18235	-6.0225	2.2409	-1.8908	-0.8505	0.7997	-0.0254	-2.0957	0.739	-0.67835	-0.1758	0.6466	0.2354	-2.1473	1.0064	-0.57045	-0.2177	0.356	0.06915	-2.1666	-0.1536	-1.1601
4	5	-17.3825	26.4991	4.5583	-63.7878	23.7828	-20.0025	-21.2619	19.9914	-0.63525	-29.4117	8.3121	-10.5498	-4.3946	16.1653	5.88535	-13.8541	15.406	0.77595	-5.4419	8.8998	1.72895	-12.7227	10.0131	-1.3548
4	13	-0.6953	1.06	0.18235	-74.9289	54.3575	-10.2857	-0.8505	0.7997	-0.0254	-33.1448	8.4369	-12.35395	-0.1758	0.6466	0.2354	-36.56	5.3276	-15.6162	-0.2177	0.356	0.06915	-25.3547	-0.7363	-13.0455
5	1	0.3228	1.962	1.1424	-9.28	-0.6248	-4.9524	-0.1624	0.4903	0.16395	-1.6191	5.6085	1.9947	-0.0936	0.3991	0.15275	-2.6479	1.591	-0.52845	-0.0836	0.2713	0.09385	-1.5331	1.525	-0.00405
5	5	8.0695	49.0509	28.5602	-56.4179	11.5407	-22.4386	-4.0595	12.257	4.09875	-21.4379	35.8279	7.195	-2.341	9.9782	3.8186	-8.9711	19.357	5.19295	-2.0909	6.7826	2.34585	-7.042	13.4871	3.22255
5	13	0.3228	1.962	1.1424	-166.1701	-8.2157	-87.1929	-0.1624	0.4903	0.16395	-22.2079	99.7154	38.75375	-0.0936	0.3991	0.15275	-59.667	30.9346	-14.3662	-0.0836	0.2713	0.09385	-37.4356	26.6751	-5.38025
6	1	1.4012	2.6102	2.0057	-10.4015	16.0763	2.8374	0.2853	0.8013	0.5433	-9.239	0.7394	-4.2498	0.0976	0.4234	0.2605	-5.8286	2.1662	-1.8312	-0.009	0.2137	0.10235	-3.3408	2.5885	-0.37615
6	5	35.0304	65.2549	50.14265	-40.9089	79.8061	19.4486	7.1329	20.0337	13.5833	-27.2903	21.6154	-2.83745	2.4408	10.5858	6.5133	-4.7223	22.0216	8.64965	-0.2254	5.3416	2.5581	-7.7198	14.43	3.3551
6	13	1.4012	2.6102	2.0057	-251.1671	316.3474	32.59015	0.2853	0.8013	0.5433	-208.8571	42.8602	-82.99845	0.0976	0.4234	0.2605	-151.366	28.2149	-61.5754	-0.009	0.2137	0.10235	-85.8293	57.8323	-13.9985
7	1	0.5214	2.528	1.5247	-3.2368	2.2405	-0.49815	0.223	1.1265	0.67475	-1.3113	2.4581	0.5734	-0.0546	0.5588	0.2521	-1.8206	0.0438	-0.8884	-0.1552	0.356	0.1004	-0.5684	0.89	0.1608
7	5	13.0358	63.201	38.1184	-24.9946	23.9255	-0.53455	5.5745	28.1627	16.8686	-6.8614	28.3751	10.75685	-1.3647	13.9697	6.3025	-9.5029	12.943	1.72005	-3.8802	8.8996	2.5097	-6.1981	5.4133	-0.3924
7	13	0.5214	2.528	1.5247	-47.946	32.8501	-7.54795	0.223	1.1265	0.67475	-22.6792	25.6623	1.49155	-0.0546	0.5588	0.2521	-20.4671	2.365	-9.05105	-0.1552	0.356	0.1004	-6.7885	15.6132	4.41235
8	1	1.56	2.6159	2.08795	-3.1278	11.1706	4.0214	0.2918	0.8345	0.56315	-1.8604	2.518	0.3288	0.0922	0.4718	0.282	-0.1984	2.4321	1.11685	-0.0184	0.2926	0.1371	-0.9475	1.6347	0.3436
8	5	38.9999	65.3963	52.1981	-21.5453	63.2599	20.8573	7.295	20.8633	14.07915	-3.6024	26.8585	11.62805	2.3053	11.7948	7.05005	-3.9343	14.5481	5.3069	-0.4591	7.3138	3.42735	-1.9989	12.7972	5.39915
8	13	1.56	2.6159	2.08795	-43.9283	225.4165	90.7441	0.2918	0.8345	0.56315	-37.367	34.361	-1.503	0.0922	0.4718	0.282	-10.3064	48.5567	19.12515	-0.0184	0.2926	0.1371	-22.1652	32.3263	5.08055
9	1	2.3723	3.3657	2.869	0.2068	21.8558	11.0313	0.5523	0.9143	0.7333	-3.0239	5.8864	1.43125	0.1888	0.4195	0.30415	-2.5849	2.3706	-0.10715	0.115	0.2718	0.1934	-2.288	1.746	-0.271
9	5	59.3066	84.1435	71.72505	7.2629	108.8884	58.07565	13.808	22.8575	18.33275	-0.4571	28.2124	13.87765	4.7209	10.4875	7.6042	-7.5198	13.6301	3.05515	2.8748	6.7958	4.8353	-4.7569	11.834	3.53855
9	13	2.3723	3.3657	2.869	-12.3367	423.3215	205.4924	0.5523	0.9143	0.7333	-73.8055	135.4485	30.8215	0.1888	0.4195	0.30415	-67.5176	51.9976	-7.76	0.115	0.2718	0.1934	-58.6795	31.6449	-13.5173

4.2.2 RM1. RM1 is a residual bootstrap method. The pseudo-experimental data is used to estimate the statistical model with which residuals are then calculated. The residuals are then bootstrapped across all experimental observations (each observation assumed independent) and new bootstrap samples are formed. Whole-plot and subplot errors are estimated for each bootstrap sample. The bootstrap sample estimates are then aggregated to form the bootstrap estimate for the 20 iterations of the 2, 5, 10 and 20 replicate experiments analyzed.

Table 4.3 compares RM1 estimates to true values. RM1 methodology does not improve the accuracy but does improve precision of the error estimates over EV. In designs with 2 replications, only 5 of the 27 subplot error and 0 of the 27 whole-plot error CIs contain the true error component. Even with 20 replications, only 5 subplot error (a different subset of 5) and 1 whole-plot error CIs contain the true error component. Even though the precision is better, the subplot error is substantially larger, while the whole-plot error is substantially smaller than the true values. Surprisingly, this method did perform better in analyzing results for Distribution 5 then did EV. It is conjectured that the distortion in the error estimates is due to the correlation between observations within the same whole-plot. If this structure is not maintained, then when the bootstrap is performed the errors will be smoothed as is indicated in the results for this method.

Bootstrapping across the residuals is not a promising method, so RM1 is not really a candidate to augment split-plot analysis. Methods that incorporate the dependence within whole-plots are examined next. The RM1 method mixes errors amon the whole-plots obscuring the estimation process thus yielding inferior estimates as compared to the EV estimates from the original pseudo-experiment.

Table 4.3: RM1 Direct Comparison Confidence Intervals

Design	Distribution	2 replications						5 replications						10 replications						20 replications					
		ΔSP			ΔWP			ΔSP			ΔWP			ΔSP			ΔWP			ΔSP			ΔWP		
		Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean
1	1	-2.2047	0.2224	-0.99115	-10.2454	-7.8779	-9.06165	0.5185	1.8956	1.20705	-7.4953	-6.0843	-6.7898	2.289	3.3594	2.8242	-5.7119	-4.6493	-5.1806	2.8483	4.1411	3.4947	-5.0776	-3.8025	-4.44005
1	5	-72.0418	-54.9686	-63.5052	-80.7141	-63.5352	-72.12465	-38.2064	-18.6874	-28.4469	-46.2546	-26.872	-36.5633	-24.5505	-3.1922	-13.87135	-32.4391	-11.1898	-21.8145	-7.1442	9.41	1.1329	-14.9566	1.3905	-6.78305
1	13	24.8329	62.8618	43.84735	-175.4934	-139.9407	-157.71705	52.0222	81.6744	66.8483	-148.2308	-118.4658	-133.3483	78.7474	96.1216	87.4345	-121.5759	-104.486	-113.031	82.5098	99.205	90.8574	-116.938	-100.499	-108.718
2	1	-0.3495	2.1022	0.87635	-16.8445	-14.6708	-15.75765	1.0059	2.7659	1.8859	-15.1749	-13.5457	-14.3603	2.5271	4.3919	3.4595	-13.6246	-11.806	-12.7153	2.9769	4.0471	3.512	-13.0715	-12.0153	-12.5434
2	5	-51.57	-30.881	-41.2255	-74.2576	-56.0561	-65.15685	-28.2893	-12.5088	-20.39905	-47.1129	-32.1527	-39.6328	-15.389	1.4916	-6.9487	-33.2406	-16.3394	-24.79	-4.7462	6.1045	0.67915	-21.1081	-10.5413	-15.8247
2	13	40.7493	78.1887	59.469	-365.8536	-334.5051	-350.17935	56.1223	86.6574	71.38985	-346.6217	-317.7936	-332.20765	83.4388	110.1709	96.80485	-317.921	-291.941	-304.931	83.5908	101.826	92.7086	-316.804	-299.276	-308.04
3	1	-0.2015	1.8068	0.80265	-33.1951	-31.6849	-32.44	2.1912	4.4418	3.3165	-30.3709	-28.2546	-29.31275	2.9664	4.0623	3.51435	-29.3097	-28.2948	-28.8023	3.3096	4.1308	3.7202	-28.8494	-28.0421	-28.4458
3	5	-35.3443	-13.2535	-24.2989	-84.7658	-67.6153	-76.19055	-8.9595	9.5552	0.29785	-48.63	-31.9997	-40.31485	-5.9904	6.2385	0.12405	-41.4127	-29.8624	-35.6376	-1.0487	6.5823	2.7668	-34.9457	-27.4308	-31.1883
3	13	30.5354	63.0098	46.7726	-778.6663	-754.5182	-766.59225	61.1774	99.5737	80.37555	-743.412	-707.4009	-725.40645	83.2482	106.012	94.6301	-720.0176	-698.387	-709.202	87.4971	107.003	97.2503	-714.111	-695.097	-704.604
4	1	-0.5376	1.9138	0.6881	-9.2087	-7.259	-8.23385	1.9858	3.1196	2.5527	-6.3823	-5.3322	-5.85725	2.7424	4.0676	3.405	-5.4507	-4.1556	-4.80315	2.8081	3.7468	3.27745	-5.2954	-4.3783	-4.83685
4	5	-51.9309	-26.8703	-39.4006	-69.4003	-49.3775	-59.3889	-27.8891	-3.9419	-15.9155	-40.0614	-17.4437	-28.75255	-8.4225	9.6873	0.6324	-18.6282	-1.496	-10.0621	-6.0225	6.9167	0.4471	-15.2716	-2.3261	-8.79885
4	13	41.1749	79.3722	60.27355	-168.4225	-137.8063	-153.1144	74.664	86.331	80.4975	-130.1477	-119.1265	-124.6371	80.8592	96.4357	88.64745	-121.2656	-105.986	-113.626	85.4333	97.5338	91.4836	-115.643	-103.767	-109.705
5	1	0.5629	1.9326	1.24775	-16.7708	-15.7505	-16.26065	2.581	4.6059	3.59345	-14.0278	-12.1279	-13.07785	2.97	4.1007	3.53535	-13.3318	-12.2486	-12.7902	3.7078	4.3216	4.0147	-12.4697	-11.863	-12.1664
5	5	-25.5269	-4.5376	-15.03225	-62.8223	-47.9664	-55.39435	-13.7761	5.9344	-3.92085	-37.3663	-19.8664	-28.61635	-4.0611	7.4872	1.71305	-24.3247	-12.851	-18.5879	0.1078	7.8436	3.9757	-18.2466	-10.6166	-14.4316
5	13	38.7121	62.3884	50.55025	-374.0072	-356.801	-365.4041	81.2896	108.8456	95.0676	-326.5175	-301.7605	-314.139	82.8788	101.4469	92.16285	-320.8215	-303.41	-312.116	91.8281	104.199	98.0133	-310.398	-298.315	-304.357
6	1	2.7351	5.3273	4.0312	-31.7032	-30.0201	-30.86165	3.0612	4.1703	3.61575	-29.8385	-28.872	-29.35525	3.348	4.2466	3.7973	-29.1147	-28.2461	-28.6804	3.5203	4.1645	3.8424	-28.6945	-28.0742	-28.3844
6	5	7.6224	29.7881	18.70525	-63.2297	-49.1821	-56.2059	0.1583	12.571	6.36465	-44.5611	-33.7298	-39.14545	1.2684	8.9962	5.1323	-37.1045	-29.4805	-33.2925	1.2431	6.703	3.97305	-33.7404	-28.3789	-31.0597
6	13	57.0426	113.7253	85.38395	-765.4566	-729.8725	-747.66455	78.9512	102.1546	90.5529	-731.5895	-711.7397	-721.6646	86.0963	106.8674	96.48185	-719.2561	-699.729	-709.493	87.9465	103.134	95.5401	-714.531	-699.704	-707.117
7	1	1.4295	3.5544	2.49195	-7.8538	-6.2209	-7.03735	2.9813	4.4966	3.73895	-5.5672	-4.1934	-4.8803	2.8416	3.7695	3.30555	-5.4017	-4.5209	-4.9613	3.5822	4.3329	3.95755	-4.5466	-3.8011	-4.17385
7	5	-22.8996	4.3981	-9.25075	-48.9278	-28.2011	-38.56445	-3.6502	16.1824	6.2661	-19.4758	-1.1607	-10.31825	-4.8382	9.1415	2.15165	-16.4673	-3.083	-9.77515	-2.062	6.6325	2.28525	-11.8782	-3.4075	-7.64285
7	13	63.3225	90.0267	76.6746	-152.6643	-132.5552	-142.60975	82.9853	99.5084	91.24685	-123.5553	-107.8447	-115.7	86.7861	97.0837	91.9349	-116.8382	-107.078	-111.958	95.8704	106.342	101.106	-106.05	-95.6931	-100.871
8	1	3.1156	5.7414	4.4285	-15.0059	-13.1697	-14.0878	3.359	4.4361	3.89755	-13.4189	-12.4383	-12.9286	3.8487	4.4891	4.1689	-12.5319	-11.9087	-12.2203	3.734	4.3127	4.02335	-12.4564	-11.888	-12.1722
8	5	8.2596	30.2451	19.25235	-40.1032	-24.5608	-32.332	2.5576	15.1129	8.83525	-24.3563	-13.112	-18.73415	2.9286	10.825	6.8768	-18.2213	-10.5799	-14.4006	2.1349	9.2103	5.6726	-16.5528	-9.6575	-13.1052
8	13	72.2047	123.4335	97.8191	-350.2416	-313.9236	-332.0826	85.0078	101.9653	93.48655	-324.4062	-309.6075	-317.00685	93.1495	106.2874	99.71845	-311.7259	-299.201	-305.463	92.664	104.886	98.7751	-309.59	-297.711	-303.65
9	1	4.9025	7.2825	6.0925	-30.0795	-28.5546	-29.31705	3.8349	4.8412	4.33805	-29.1631	-28.2618	-28.71245	3.7616	4.3834	4.0725	-28.7118	-28.1209	-28.4164	3.7238	4.3175	4.02065	-28.5121	-27.9381	-28.2251
9	5	36.7995	62.2058	49.50265	-41.5264	-24.5337	-33.03005	10.4331	19.2101	14.8216	-35.4903	-27.4874	-31.48885	5.2269	10.9346	8.08075	-32.8594	-27.5989	-30.2292	4.9564	9.0025	6.97945	-30.0864	-26.1414	-28.1139
9	13	86.1132	126.6652	106.3892	-744.4708	-717.9288	-731.1998	88.7576	112.2475	100.50255	-722.6252	-702.1693	-712.39725	90.0767	104.253	97.16485	-715.6965	-702.301	-708.999	91.1848	103.233	97.2087	-711.585	-699.978	-705.781

4.2.3 RM2. RM2 is a residual bootstrap method devised to address the correlation between observations within whole-plots. In RM2, residuals are re-sampled within a whole-plot. New bootstrap samples are formed. Whole-plot and subplot errors are estimated for each bootstrap sample. The bootstrap sample estimates are then aggregated to form the bootstrap estimate for the 20 iterations of the 2, 5, 10 and 20 replicate experiments analyzed.

Table 4.4 compares RM2 estimates to true values. For designs with lower replications, the RM2 methodology improves (in accuracy) over RM1 estimates and performs as well, or better, than EV in many cases. However, as experimental replication increases, the accuracy improvement over EV disappears until there is no improvement in the precision of the whole-plot error estimate. The precision of the subplot error estimates improve, but the estimates are biased low. In designs with 2 replications, 6 of the 27 subplot error and 21 of the 27 whole-plot error confidence intervals contain the true error component. In designs with 20 replications, 0 subplot error and 11 whole-plot error confidence intervals contain the true error component. RM2 may be sufficient to augment EV in providing more accurate subplot error estimates with better precision for Designs 6, 8, and 9 throughout the spectrum of distribution sets analyzed in this research for experiments between 2 and 5 replications even though accuracy and precision is not improved for whole-plot error estimates. Further investigation for these particular experiments may be needed. Further investigation on methods that account for the correlation within whole-plots using other bootstrap techniques is another avenue of further investigation. While the whole-plot sampling seems more intuitive, the sampling seems to distort the subplot error estimate and thus the whole-plot error estimate.

Table 4.4: RM2 Direct Comparison Confidence Intervals

Design	Distribution	2 replications						5 replications					
		ΔSP			ΔWP			ΔSP			ΔWP		
		Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean
1	1	-3.7116	-3.2534	-3.4825	-8.3112	6.6367	-0.83725	-2.8388	-2.3192	-2.579	-3.1322	1.0946	-1.0188
1	5	-92.6579	-81.2582	-86.95805	-32.7478	91.8835	29.56785	-71.0692	-58.1309	-64.60005	-6.1555	63.6479	28.7462
1	13	-3.708	-3.2794	-3.4937	-131.7085	138.0359	3.1637	-2.8388	-2.3071	-2.57295	-82.5292	0.0803	-41.22445
2	1	-2.6296	-1.8595	-2.24455	-12.1437	4.1468	-3.99845	-1.9516	-1.4496	-1.7006	-5.1409	9.146	2.00255
2	5	-65.6738	-46.4139	-56.04385	-42.4099	64.1786	10.88435	-48.61	-36.1425	-42.37625	22.3758	93.7443	58.06005
2	13	-2.6294	-1.8676	-2.2485	-228.3457	33.8725	-97.2366	-1.9527	-1.4494	-1.70105	-135.7883	100.1746	-17.80685
3	1	-1.6498	-0.8072	-1.2285	-23.5362	9.2202	-7.158	-0.9425	-0.4145	-0.6785	-10.1514	7.5074	-1.322
3	5	-41.2452	-20.1808	-30.713	-50.6278	21.5421	-14.54285	-23.5623	-10.3629	-16.9626	44.8409	134.5625	89.7017
3	13	-1.6498	-0.8072	-1.2285	-529.6548	199.1353	-165.25975	-0.9425	-0.4145	-0.6785	-352.8534	-2.6844	-177.7689
4	1	-2.9711	-2.4258	-2.69845	-4.9113	3.3355	-0.7879	-2.6635	-1.9696	-2.31655	-0.4865	2.2722	0.89285
4	5	-74.2771	-60.6451	-67.4611	-38.4706	52.3148	6.9221	-66.5884	-49.2404	-57.9144	8.3606	49.6121	28.98635
4	13	-2.9697	-2.4213	-2.6955	-73.8597	55.3373	-9.2612	-2.6546	-1.949	-2.3018	-31.5447	9.9709	-10.7869
5	1	-1.707	-0.8217	-1.26435	-7.2153	1.24	-2.98765	-1.4621	-1.0337	-1.2479	0.8009	8.123	4.46195
5	5	-42.6745	-20.5419	-31.6082	-7.1774	59.8341	26.32835	-36.5515	-25.8428	-31.19715	37.6957	101.2524	69.47405
5	13	-1.7026	-0.8218	-1.2622	-164.3251	-6.8426	-85.58385	-1.458	-1.0251	-1.24155	-19.6493	102.2933	41.322
6	1	-0.2795	0.546	0.13325	-7.6252	18.525	5.4499	-0.569	-0.1601	-0.36455	-6.1838	3.9105	-1.13665
6	5	-6.9873	13.6491	3.3309	26.1159	143.2597	84.6878	-14.225	-4.0022	-9.1136	49.7159	102.8883	76.3021
6	13	-0.2854	0.5456	0.1301	-248.5977	318.8827	35.1425	-0.5737	-0.1596	-0.36665	-205.5739	46.0092	-79.78235
7	1	-2.3024	-1.5482	-1.9253	-1.7735	3.9227	1.0746	-2.097	-1.6975	-1.89725	0.5759	4.425	2.50045
7	5	-57.5596	-38.706	-48.1328	11.8139	65.2673	38.5406	-52.4251	-42.4386	-47.43185	39.6373	78.4675	59.0524
7	13	-2.3068	-1.5394	-1.9231	-46.3762	34.6619	-5.85715	-2.1003	-1.6823	-1.8913	-20.7027	27.6474	3.47235
8	1	-0.6132	0.0344	-0.2894	-0.4676	13.7532	6.6428	-1.0233	-0.6455	-0.8344	0.9771	5.3883	3.1827
8	5	-15.212	0.8914	-7.1603	44.0835	129.4062	86.74485	-25.5951	-16.0552	-20.82515	65.4248	99.4023	82.41355
8	13	-0.6064	0.0295	-0.28845	-41.3362	227.8304	93.2471	-1.0206	-0.6439	-0.83225	-34.5832	37.1478	1.2823
9	1	0.8301	1.5745	1.2023	3.6976	25.4719	14.58475	-0.2313	0.07	-0.08065	0.428	9.4319	4.92995
9	5	20.5291	39.4645	29.9968	88.8367	198.0867	143.4617	-5.7942	1.7569	-2.01865	84.453	115.4678	99.9604
9	13	0.8283	1.5857	1.207	-9.3588	427.1271	208.88415	-0.2288	0.0707	-0.07905	-70.1057	138.9063	34.4003

Design	Distribution	10 replications						20 replications					
		ΔSP			ΔWP			ΔSP			ΔWP		
		Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean
1	1	-2.6518	-1.9848	-2.3183	-0.8542	1.8805	0.51315	-2.2239	-1.8528	-2.03835	-0.8056	2.8178	1.0061
1	5	-66.2221	-49.1153	-57.6687	9.7456	44.6921	27.21885	-55.3883	-46.1899	-50.7891	35.8603	67.1712	51.51575
1	13	-2.656	-1.9768	-2.3164	-22.9987	21.6862	-0.65625	-2.2207	-1.856	-2.03835	-49.376	4.9438	-22.2161
2	1	-1.5551	-1.179	-1.36705	-0.6452	6.7465	3.05065	-1.2978	-0.969	-1.1334	0.2819	4.3642	2.32305
2	5	-38.8196	-29.5545	-34.18705	33.1242	94.9202	64.0222	-32.5041	-24.1492	-28.3267	57.5707	90.0415	73.8061
2	13	-1.5492	-1.1782	-1.3637	-47.6918	136.9687	44.63845	-1.2995	-0.9685	-1.134	-57.8658	40.0287	-8.91855
3	1	-0.8171	-0.4081	-0.6126	-6.2229	7.753	0.76505	-0.662	-0.3977	-0.52985	-4.0508	4.6431	0.29615
3	5	-20.428	-10.2013	-15.31465	47.4834	107.9414	77.7124	-16.5499	-9.942	-13.246	76.636	111.5783	94.10715
3	13	-0.8171	-0.4081	-0.6126	-160.1126	146.273	-6.9198	-0.662	-0.3977	-0.52985	-192.921	5.6654	-93.628
4	1	-2.2346	-1.8476	-2.0411	-0.2747	2.927	1.32615	-2.1811	-1.9039	-2.0425	-0.2781	1.8245	0.7732
4	5	-55.8651	-46.1902	-51.02765	32.8397	63.8623	48.351	-54.5268	-47.598	-51.0624	34.288	59.5607	46.92435
4	13	-2.2336	-1.8521	-2.04285	-34.5777	7.3096	-13.6341	-2.1789	-1.9015	-2.0402	-23.4007	1.2689	-11.0659
5	1	-1.2419	-0.891	-1.06645	0.0841	4.4045	2.2443	-1.1486	-0.8916	-1.0201	1.3892	4.4538	2.9215
5	5	-31.0479	-22.2755	-26.6617	58.9744	90.6099	74.79215	-28.7143	-22.2898	-25.5021	65.212	87.0456	76.1288
5	13	-1.2448	-0.9003	-1.07255	-56.9539	33.752	-11.601	-1.1488	-0.8891	-1.01895	-34.4293	29.6534	-2.38795
6	1	-0.5679	-0.2979	-0.4329	-2.4721	5.5749	1.5514	-0.5804	-0.3905	-0.48545	0.0708	6.0249	3.04785
6	5	-14.1973	-7.4475	-10.8224	78.5736	107.2341	92.90385	-14.5102	-9.7616	-12.1359	76.992	100.4297	88.71085
6	13	-0.5708	-0.2963	-0.43355	-147.8331	31.7766	-58.0283	-0.5824	-0.3925	-0.48745	-82.3749	61.2701	-10.5524
7	1	-2.1172	-1.8265	-1.97185	0.0969	1.9941	1.0455	-2.1225	-1.875	-1.99875	1.3347	2.9044	2.11955
7	5	-52.93	-45.6616	-49.2958	37.4827	62.8815	50.1821	-53.0631	-46.8752	-49.9692	41.7539	55.6035	48.6787
7	13	-2.1288	-1.8379	-1.98335	-18.4854	4.2881	-7.09865	-2.1257	-1.8738	-1.99975	-4.8457	17.6836	6.41895
8	1	-1.0463	-0.7737	-0.91	2.6993	5.3791	4.0392	-1.0724	-0.8437	-0.95805	2.0258	4.5861	3.30595
8	5	-26.1771	-19.3308	-22.75395	69.1951	88.9578	79.07645	-26.7226	-20.9853	-23.854	71.2146	87.7178	79.4662
8	13	-1.045	-0.7734	-0.9092	-7.5752	51.4517	21.93825	-1.0711	-0.8395	-0.9553	-19.2347	35.2139	7.9896
9	1	-0.4365	-0.2392	-0.33785	0.8467	5.8205	3.3336	-0.4498	-0.3137	-0.38175	1.2106	5.2831	3.24685
9	5	-10.8284	-5.9297	-8.37905	78.0205	100.4685	89.2445	-11.1467	-7.8115	-9.4791	82.4286	99.782	91.1053
9	13	-0.4339	-0.2368	-0.33535	-64.1145	55.3175	-4.3985	-0.4468	-0.3112	-0.379	-55.1704	35.2453	-9.96255

4.2.4 RM3. RM3 is the final residual bootstrap method discussed in this research. RM3 separately re-samples the whole-plot and subplot residuals. The two bootstrapped residual types form the error term as in Equation 3.1. With the new bootstrap samples, whole-plot and subplot errors are estimated for each bootstrap sample. The bootstrap sample estimates are then aggregated to form the bootstrap estimate and 20 iterations of 2, 5, 10 and 20 replicate experiments are analyzed.

Table 4.5compares RM3 estimates to true values. Splitting the residuals into whole-plot and subplot residuals and bootstrapping both residual types do not appear as effective, but provides better precision in the estimates. In designs with 2 replications, 5 of the 27 subplot error and 1 of the 27 whole-plot error confidence intervals contain the true error component. In designs with 20 replications, 0 subplot error and 12 whole-plot error confidence intervals contain the true error component. RM3 does show promise in improving subplot error estimate accuracy and precision for designs at least the size of Designs 8 and 9. However, further improvements in whole-plot error estimation accuracy is highly unlikely. The results may improve in accuracy if the whole-plot residuals are re-sampled without replacement; this approach was not examined in this research. In general, this further delineation of re-sampling, down to both the subplot and whole-plot level is not providing improved precision in the error component estimates.

Table 4.5: RM3 Direct Comparison Confidence Intervals

Design	Distribution	2 replications						5 replications						10 replications						20 replications					
		ΔSP			ΔWP			ΔSP			ΔWP			ΔSP			ΔWP			ΔSP			ΔWP		
		Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean
1	1	-3.7090	-3.2419	-3.4755	-8.9921	-4.1309	-6.5615	-2.8383	-2.3108	-2.5746	-4.2564	-1.6533	-2.9548	-2.6536	-1.9735	-2.3135	-0.9697	1.0243	0.0273	-2.2189	-1.8467	-2.0328	-0.0999	2.3631	1.1316
1	5	-92.7802	-81.1322	-86.9562	-64.2184	-30.9255	-47.5720	-71.0117	-57.9614	-64.4866	-15.4390	17.5973	1.0791	-66.1487	-49.1151	-57.6319	6.4882	37.7697	22.1289	-55.5883	-46.3569	-50.9726	32.2488	59.7681	46.0085
1	13	-3.7066	-3.2468	-3.4767	-146.1026	-69.5467	-107.8246	-2.8424	-2.3338	-2.5881	-93.6257	-35.6638	-64.6448	-2.3307	-1.8129	-2.0718	-42.5066	-7.2333	-24.8699	-2.2232	-1.8570	-2.0401	-32.1925	0.8160	-15.6883
2	1	-2.6337	-1.8541	-2.2439	-11.8782	-3.8173	-7.8477	-1.9457	-1.4449	-1.6953	-7.3425	-0.6121	-3.9773	-1.5518	-1.1761	-1.3640	-1.9007	4.8311	1.4652	-1.2967	-0.9668	-1.1318	-0.6608	3.3225	1.3309
2	5	-65.7875	-46.6144	-56.2010	-47.0571	-9.8558	-28.4565	-48.7739	-36.3618	-42.5678	6.3902	40.7784	23.5843	-38.7571	-29.4517	-34.1044	34.0888	79.5522	56.8205	-32.4581	-24.1416	-28.2999	57.2773	83.6149	70.4461
2	13	-2.6319	-1.8481	-2.2400	-254.5432	-124.7845	-189.6639	-1.9510	-1.4496	-1.7003	-178.2761	-57.2993	-117.7877	-1.5436	-1.1742	-1.3589	-67.8607	37.3482	-15.2562	-1.3053	-0.9712	-1.1383	-64.1971	7.4759	-28.3606
3	1	-1.6531	-0.8094	-1.2312	-26.5475	-17.2911	-21.9193	-0.9373	-0.4165	-0.6769	-11.0545	4.3371	-3.3587	-0.8170	-0.4021	-0.6096	-4.6574	2.5391	-1.0592	-0.6616	-0.3969	-0.5293	-2.1030	3.8030	0.8500
3	5	-41.4402	-20.3793	-30.9098	-59.2264	-27.7234	-43.4749	-23.4290	-10.4799	-16.9544	37.3446	107.5242	72.4344	-20.2979	-10.1430	-15.2204	51.5603	86.4764	69.0183	-16.4712	-9.8508	-13.1610	67.1522	89.8943	78.5233
3	13	-1.6495	-0.8033	-1.2264	-612.0091	-420.4795	-516.2443	-0.9402	-0.4146	-0.6774	-347.0908	-67.4824	-207.2866	-0.8145	-0.4027	-0.6086	-155.6605	19.2913	-68.1846	-0.6583	-0.3964	-0.5273	-108.8851	46.4555	-31.2148
4	1	-1.9530	0.0446	-0.9542	-8.9072	-6.6902	-7.7987	-0.4359	0.4141	-0.0109	-5.6702	-4.5041	-5.0872	0.0768	1.0426	0.5597	-4.6132	-3.1091	-3.8612	0.2788	0.8199	0.5493	-4.4974	-3.3578	-3.9276
4	5	-63.6845	-40.6314	-52.1579	-66.9928	-44.9324	-55.9626	-47.0721	-29.0591	-38.0656	-33.9651	-9.5330	-21.7490	-31.7814	-18.8705	-25.3259	-11.5576	7.9749	-1.7914	-30.7819	-20.5991	-25.6905	-7.3185	6.8508	-0.2339
4	13	17.4719	43.2943	30.3831	-163.3446	-127.6812	-145.5129	35.2072	44.0288	39.6180	-118.3039	-105.2018	-111.7529	34.0363	47.2611	40.6487	-107.2466	-88.9169	-98.0817	41.0989	46.2074	43.6531	-101.6875	-86.0099	-93.8487
5	1	-0.2638	0.7795	0.2578	-15.9724	-14.3021	-15.1373	0.6199	1.5641	1.0920	-11.3475	-7.4611	-9.4043	0.9322	1.5798	1.2560	-10.3572	-7.9982	-9.1777	1.1599	1.6579	1.4089	-8.5342	-7.2415	-7.8879
5	5	-33.3316	-11.0835	-22.2075	-56.3647	-39.8708	-48.1177	-25.6911	-9.9294	-17.8103	-21.1183	5.5630	-7.7777	-17.5366	-7.4034	-12.4700	-3.4951	12.4375	4.4712	-15.2161	-7.9035	-11.5598	5.9171	16.2218	11.0694
5	13	20.4471	38.9709	29.7090	-356.6474	-325.4470	-341.0472	38.2880	57.7782	48.0331	-271.3459	-213.6786	-242.5122	38.7217	51.9968	45.3592	-259.4890	-217.1800	-238.3345	44.2511	52.6354	48.4433	-237.7153	-208.8187	-223.2670
6	1	1.5923	3.3177	2.4550	-29.5850	-23.4088	-26.4969	1.0626	2.0081	1.5353	-23.1854	-19.8119	-21.4986	1.2261	1.9007	1.5634	-21.0774	-18.1392	-19.6083	1.4489	1.9010	1.6750	-20.7677	-17.8049	-19.2863
6	5	0.8147	22.4730	11.6438	-50.7165	-25.6587	-38.1876	-7.7273	4.7019	-1.5127	-16.6553	-3.0379	-9.8466	-6.1048	1.5650	-2.2699	-8.2860	1.0875	-3.5992	-6.3380	-1.1054	-3.7217	-3.7721	7.0827	1.6553
6	13	29.5259	61.8494	45.6876	-702.0443	-571.2108	-636.6276	35.4570	51.2413	43.3491	-585.1248	-499.0823	-542.1035	36.2727	48.9935	42.6331	-532.0168	-447.4232	-489.7200	41.4964	51.5526	46.5245	-536.2090	-467.4431	-501.8260
7	1	-2.3052	-1.5474	-1.9263	-5.1529	-2.1853	-3.6691	-2.0993	-1.6904	-1.8948	-1.0041	1.6506	0.3232	-2.1287	-1.8375	-1.9831	-0.7074	0.9681	0.1303	-2.1249	-1.8759	-2.0004	0.9958	2.3540	1.6749
7	5	-57.4919	-38.6222	-48.0571	-23.9195	8.5573	-7.6811	-52.7118	-42.4641	-47.5880	23.4675	54.2090	38.8382	-53.1197	-45.8337	-49.4767	28.8659	51.2164	40.0412	-53.1753	-46.9866	-50.0809	37.2480	50.0474	43.6477
7	13	-2.3068	-1.5505	-1.9287	-103.6590	-63.4383	-83.5487	-2.0958	-1.6852	-1.8905	-45.5796	-15.2958	-30.4377	-2.1259	-1.8336	-1.9798	-31.1804	-11.3796	-21.2800	-2.1267	-1.8753	-2.0010	-10.0640	10.5598	0.2479
8	1	-0.6099	0.0291	-0.2904	-7.8126	-0.3613	-4.0869	-1.0235	-0.6493	-0.8364	-2.0999	1.5141	-0.2929	-1.0493	-0.7780	-0.9136	1.1068	3.2893	2.1981	-1.0716	-0.8431	-0.9574	1.2692	3.5417	2.4055
8	5	-15.2066	0.9963	-7.1052	2.2553	46.1943	24.2248	-25.5947	-16.1262	-20.8605	46.4822	73.8007	60.1415	-26.1991	-19.2499	-22.7245	60.8300	78.9527	69.8914	-26.8236	-21.0543	-23.9389	65.9125	81.2525	73.5825
8	13	-0.6083	0.0293	-0.2895	-196.9807	-54.1969	-125.5888	-1.0294	-0.6483	-0.8389	-93.0929	-33.0782	-63.0855	-1.0461	-0.7709	-0.9085	-43.3825	5.9072	-18.7376	-1.0709	-0.8402	-0.9555	-35.7902	12.1133	-11.8384
9	1	0.8203	1.5678	1.1940	-12.1557	-1.2273	-6.6915	-0.2301	0.0699	-0.0801	-4.8143	1.6553	-1.5795	-0.4301	-0.2335	-0.3318	-1.6820	2.9043	0.6112	-0.4488	-0.3132	-0.3810	-0.5589	3.6654	1.5533
9	5	20.9578	39.9649	30.4614	26.0282	85.4718	55.7500	-5.7423	1.7870	-1.9776	59.6459	83.5882	71.6171	-10.7888	-5.9045	-8.3467	67.8342	88.3202	78.0772	-11.1896	-7.7731	-9.4813	76.2076	91.3881	83.7979
9	13	0.8225	1.5766	1.1996	-352.3275	-137.8643	-245.0959	-0.2288	0.0664	-0.0812	-178.7369	-13.8667	-96.3018	-0.4320	-0.2353	-0.3337	-121.0885	-14.5501	-67.8193	-0.4468	-0.3099	-0.3783	-89.3352	4.4246	-42.4553

4.2.5 OM1. OM1 is an observational bootstrap method that re-samples across observational replicates. New bootstrap samples are formed and whole-plot and subplot errors are estimated for each bootstrap sample. The bootstrap sample estimates are then aggregated to form the bootstrap estimate. 20 iterations of 2, 5, 10 and 20 replicate experiments are analyzed.

Table 4.6 compares 0M1 estimates to the true values. The mean difference and CI widths are similar to those from RM1. In designs with 2 replications, 5 of the 27 subplot error and 0 of the 27 whole-plot error confidence intervals contain the true error component. In designs with 20 replications, only 5 subplot error and 1 whole-plot error confidence intervals contain the true error component. Surprisingly again, OM1 did improve EV estimates for distribution 5. While this re-sampling method should have been quite viable, particularly with highly replicated pseudo-experiments, the method actually performed quite poorly.

Table 4.6: OM1 Direct Comparison Confidence Intervals

Design	Distribution	2 replications						5 replications						10 replications						20 replications					
		ΔSP			ΔWP			ΔSP			ΔWP			ΔSP			ΔWP			ΔSP			ΔWP		
		Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean
1	1	-2.2191	0.0545	-1.0823	-10.1969	-7.8576	-9.0273	0.5261	1.9260	1.2261	-7.4844	-6.0681	-6.7762	2.2702	3.3505	2.8104	-5.7374	-4.6673	-5.2023	2.8641	4.1331	3.4986	-5.1280	-3.8389	-4.4835
1	5	-72.1341	-55.0121	-63.5731	-80.0845	-63.2858	-71.6852	-38.4522	-18.7409	-28.5965	-46.1941	-26.0166	-36.1054	-24.5726	-3.6325	-14.1025	-32.3727	-11.3836	-21.8782	-7.1184	9.6506	1.2661	-15.0639	1.7083	-6.6778
1	13	25.0772	61.4662	43.2717	-174.6638	-135.4608	-155.0623	51.7780	81.2280	66.5030	-148.3452	-119.6443	-133.9948	78.2945	96.2784	87.2864	-121.6165	-104.4150	-113.0158	82.6454	99.1901	90.9178	-117.3441	-100.9830	-109.1635
2	1	-0.0148	2.5317	1.2584	-17.1517	-15.0228	-16.0872	1.0726	2.8224	1.9475	-15.2017	-13.5165	-14.3591	2.5427	4.3727	3.4577	-13.5783	-11.7828	-12.6806	2.9885	4.0646	3.5266	-13.0746	-12.0416	-12.5581
2	5	-43.1620	-19.2050	-31.1835	-81.3609	-64.1315	-72.7462	-27.2275	-11.7558	-19.4917	-47.8994	-32.5841	-40.2417	-15.0821	1.7436	-6.6692	-33.0009	-16.2781	-24.6395	-4.5148	6.2723	0.8788	-21.9590	-11.2976	-16.6283
2	13	41.1102	78.4021	59.7562	-366.5825	-334.6758	-350.6292	56.0463	86.8754	71.4608	-346.2452	-317.6087	-331.9269	83.3756	110.1419	96.7587	-318.3531	-292.9205	-305.6368	83.6171	101.8480	92.7325	-317.0960	-298.9844	-308.0402
3	1	0.2380	2.3621	1.3000	-33.6787	-32.2920	-32.9854	2.2748	4.5410	3.4079	-30.5825	-28.4528	-29.5176	2.9845	4.0813	3.5329	-29.4452	-28.4092	-28.9272	3.3236	4.1412	3.7324	-28.8828	-28.0500	-28.4664
3	5	-25.7424	1.4542	-12.1441	-94.9776	-83.2574	-89.1175	-6.8361	12.1260	2.6450	-55.6666	-39.5065	-47.5866	-5.2239	7.2320	1.0040	-43.3398	-32.3263	-37.8331	-0.8359	6.8141	2.9891	-36.0212	-28.3725	-32.1969
3	13	30.8262	63.4520	47.1391	-779.3320	-756.1190	-767.7255	61.4460	99.8342	80.6401	-744.7501	-709.5545	-727.1523	83.1271	105.9657	94.5464	-720.1853	-698.4938	-709.3396	87.5635	107.0266	97.2950	-714.0383	-694.6857	-704.3620
4	1	-0.1818	2.2644	1.0413	-9.3911	-7.4333	-8.4122	2.0191	3.2534	2.6363	-6.4174	-5.4441	-5.9307	2.7736	4.1400	3.4568	-5.4730	-4.1911	-4.8320	2.8233	3.7700	3.2967	-5.3132	-4.3867	-4.8499
4	5	-43.8795	-16.4566	-30.1680	-73.3887	-52.1959	-62.7923	-27.1914	0.3471	-13.4221	-40.5105	-19.5701	-30.0403	-7.3234	11.3057	1.9911	-19.5234	-2.4128	-10.9681	-5.5779	7.7967	1.1094	-15.6509	-2.9537	-9.3023
4	13	41.5624	78.8102	60.1863	-168.3257	-137.8334	-153.0795	74.3436	86.1942	80.2689	-130.3259	-118.9163	-124.6211	81.0215	96.8183	88.9199	-121.3251	-106.0855	-113.7053	85.2813	97.3102	91.2958	-115.6956	-103.9616	-109.8286
5	1	1.2122	2.7317	1.9720	-17.3249	-16.3497	-16.8373	2.7011	4.7516	3.7264	-14.2419	-12.4489	-13.3454	3.0246	4.1650	3.5948	-13.4695	-12.4015	-12.9355	3.7251	4.3396	4.0324	-12.5144	-11.9220	-12.2182
5	5	-10.1969	17.0807	3.4419	-76.1539	-62.8086	-69.4812	-10.9172	10.1447	-0.3862	-42.9758	-27.5692	-35.2725	-2.6198	9.2184	3.2993	-27.5254	-16.4149	-21.9702	0.6081	8.5209	4.5645	-19.4713	-12.0569	-15.7641
5	13	39.5553	63.3247	51.4400	-374.3143	-357.4755	-365.8949	81.1572	108.6956	94.9264	-327.0040	-302.6042	-314.8041	82.8330	101.3260	92.0795	-321.2941	-303.4942	-312.3941	91.7146	104.2897	98.0022	-310.3118	-298.1844	-304.2481
6	1	3.3403	6.0652	4.7028	-32.6228	-31.0749	-31.8488	3.1754	4.2863	3.7308	-30.2402	-29.2571	-29.7486	3.3891	4.2816	3.8353	-29.3022	-28.4375	-28.8699	3.5384	4.1857	3.8620	-28.7771	-28.1723	-28.4747
6	5	21.6811	49.0511	35.3661	-85.3711	-75.4510	-80.4111	2.5770	15.5487	9.0628	-54.3863	-44.8756	-49.6310	2.2296	9.9712	6.1004	-41.9167	-34.3145	-38.1156	1.5364	7.0254	4.2809	-35.8645	-30.6721	-33.2683
6	13	57.6569	114.2418	85.9493	-765.9600	-729.4834	-747.7217	78.7905	102.2702	90.5303	-731.8879	-711.3995	-721.6437	86.2365	106.9281	96.5823	-719.5108	-700.1293	-709.8201	87.9194	103.1612	95.5403	-714.8354	-699.9954	-707.4154
7	1	2.0170	4.8112	3.4141	-7.7119	-5.5178	-6.6149	3.2517	4.8214	4.0366	-5.7396	-4.3685	-5.0540	2.9077	3.8258	3.3668	-5.4319	-4.5498	-4.9909	3.6119	4.3401	3.9760	-4.5745	-3.7942	-4.1844
7	5	0.0093	42.6108	21.3100	-56.3020	-31.4794	-43.8907	1.8596	24.4614	13.1605	-23.5167	-6.1276	-14.8222	-3.1575	11.3460	4.0943	-18.0511	-4.5778	-11.3145	-0.9473	7.2270	3.1398	-12.7580	-3.8280	-8.2930
7	13	56.8572	82.4174	69.6373	-142.0882	-106.2580	-124.1731	82.4179	99.2412	90.8296	-122.8832	-107.4071	-115.1452	86.0968	96.4053	91.2511	-116.0013	-105.8780	-110.9396	95.2453	105.7572	100.5013	-105.2811	-94.9844	-100.1328
8	1	3.5682	6.2071	4.8876	-15.2023	-11.4526	-13.3275	3.3912	4.5296	3.9604	-13.3054	-12.2373	-12.7713	3.8621	4.5143	4.1882	-12.5410	-11.8299	-12.1855	3.7416	4.3133	4.0275	-12.4564	-11.8183	-12.1374
8	5	23.4324	51.7281	37.5802	-53.0796	-41.1194	-47.0995	6.0328	18.5754	12.3041	-31.8121	-17.2403	-24.5262	4.5226	12.4704	8.4965	-22.0548	-12.8835	-17.4691	2.6739	9.6608	6.1673	-18.0222	-11.0430	-14.5326
8	13	64.7457	113.5405	89.1431	-327.5962	-243.5005	-285.5484	81.9559	100.6607	91.3083	-315.6799	-300.7183	-308.1991	92.3645	105.4270	98.8958	-309.0103	-295.7529	-302.3816	92.1832	104.1009	98.1420	-307.3973	-294.3068	-300.8521
9	1	5.0919	7.5932	6.3426	-28.7467	-23.5390	-26.1428	3.8804	4.8502	4.3653	-29.4272	-27.5870	-28.5071	3.7576	4.3691	4.0633	-28.6528	-27.6890	-28.1709	3.7204	4.3132	4.0168	-28.4837	-27.8924	-28.1881
9	5	51.5153	78.8473	65.1813	-62.8105	-24.0600	-43.4352	12.8078	21.5055	17.1567	-45.0642	-35.3956	-40.2299	6.1915	12.0650	9.1282	-37.3663	-31.8251	-34.5957	5.3093	9.3706	7.3399	-32.2170	-28.2813	-30.2492
9	13	76.5847	119.2791	97.9319	-695.8298	-585.7940	-640.8119	87.2749	110.2370	98.7560	-716.3442	-677.3387	-696.8414	89.0768	102.8957	95.9863	-709.8987	-688.2980	-699.0984	90.8538	102.9986	96.9262	-708.5505	-697.0586	-702.8045

4.2.6 OM2. OM2 is an observational bootstrap method that re-samples across observational replicates similar to OM1. However, this method expands the 2, 5, 10, 20 replication designs to 250 replication designs. New bootstrap samples are formed and whole-plot and subplot errors are estimated for each bootstrap sample. The bootstrap sample estimates are then aggregated to form the bootstrap estimate with 20 iterations of 2, 5, 10 and 20 replicate experiments are analyzed.

Table 4.7 compares OM2 estimates to the true values. The mean difference and CI widths are similar to RM1 and OM1 results. In addition, with designs with 2 replications, 6 of the 27 subplot error and 0 of the 27 whole-plot error confidence intervals contain the true error component. In designs with 20 replications, 8 subplot error and 1 whole-plot error confidence intervals contain the true error component. As with OM1, OM2 improved EV results for distribution 5 but even with the large increase in sample size the re-sampling method is not yielding improved error component estimates.

Table 4.7: OM2 Direct Comparison Confidence Intervals

Design	Distribution	2 replications						5 replications						10 replications						20 replications					
		ΔSP			ΔWP			ΔSP			ΔWP			ΔSP			ΔWP			ΔSP			ΔWP		
		Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean	Lower Bound	Upper Bound	Mean
1	1	-2.2578	0.1563	-1.0508	-10.2082	-7.9940	-9.1011	0.4569	1.8956	1.1763	-7.4927	-6.0514	-6.7720	2.3885	3.2143	2.8014	-6.0009	-4.6468	-5.3239	2.8504	4.1080	3.4792	-5.0596	-3.6878	-4.3737
1	5	-72.4122	-54.4827	-63.4474	-79.3599	-61.9864	-70.6731	-38.3388	-19.0998	-28.7193	-47.2752	-26.9393	-37.1073	-25.7306	-6.5246	-16.1276	-30.8836	-10.5069	-20.6953	-7.5158	11.6635	2.0739	-17.4059	1.3629	-8.0215
1	13	24.8157	62.9541	43.8849	-174.7199	-138.8348	-156.7774	51.2938	81.2354	66.2646	-149.5875	-119.5643	-134.5759	77.3337	93.9251	85.6294	-122.2512	-102.0360	-112.1436	82.0897	101.3292	91.7095	-116.9223	-100.7198	-108.8211
2	1	-0.7550	1.4142	0.3296	-17.2336	-15.2022	-16.2179	0.7758	2.4745	1.6252	-15.3893	-13.6924	-14.5408	2.4420	4.2940	3.3680	-13.6518	-11.5990	-12.6254	2.8858	3.9574	3.4216	-13.4138	-12.1957	-12.8048
2	5	-56.1688	-37.8689	-47.0188	-81.9370	-63.5375	-72.7372	-30.7077	-15.9075	-23.3076	-48.1781	-32.2917	-40.2349	-16.6018	-0.0125	-8.3071	-33.7102	-15.6739	-24.6920	-6.5443	5.8342	-0.3550	-20.0860	-8.1740	-14.1300
2	13	34.4636	66.8242	50.6439	-365.8866	-335.2567	-350.5717	53.3066	84.2881	68.7973	-345.2146	-317.5302	-331.3724	81.8942	107.5414	94.7178	-320.5654	-289.2679	-304.9167	84.7454	100.9469	92.8461	-316.9320	-299.0791	-308.0056
3	1	-1.1238	0.3639	-0.3799	-33.6541	-32.3716	-33.0128	1.7395	3.7682	2.7538	-30.8216	-28.5878	-29.7047	2.7617	3.8054	3.2836	-29.4118	-28.4361	-28.9239	3.1381	3.9934	3.5657	-28.9076	-27.6357	-28.2717
3	5	-50.8016	-33.7690	-42.2853	-95.8727	-83.9943	-89.9335	-15.5701	0.7513	-7.4094	-57.2378	-37.0784	-47.1581	-8.4451	2.8527	-2.7962	-47.0411	-29.3429	-38.1920	-2.0453	6.1781	2.0664	-40.2168	-26.5411	-33.3790
3	13	21.4381	45.0311	33.2346	-778.3747	-752.7964	-765.5855	55.6129	90.6359	73.1244	-744.1159	-707.5102	-725.8130	80.3115	101.9761	91.1438	-719.1530	-695.1969	-707.1749	85.4252	104.8018	95.1135	-711.9448	-693.3463	-702.6455
4	1	-1.0969	0.8691	-0.1139	-9.3755	-7.5020	-8.4388	1.6062	2.7205	2.1634	-6.3450	-5.3844	-5.8647	2.5770	4.0820	3.3295	-5.5176	-4.0850	-4.8013	2.7243	3.6749	3.1996	-5.4663	-4.3317	-4.8990
4	5	-58.0889	-37.1933	-47.6411	-73.1180	-52.6231	-62.8706	-30.0886	-3.6916	-16.8901	-40.9131	-20.9819	-30.9475	-10.2520	8.3091	-0.9714	-19.5608	-1.9941	-10.7774	-8.5347	5.4728	-1.5310	-17.5594	-4.2034	-10.8814
4	13	32.2472	62.5695	47.4083	-168.3893	-138.1375	-153.2634	69.8934	81.9515	75.9225	-130.2993	-118.8843	-124.5918	78.7789	92.8940	85.8364	-122.9684	-107.7161	-115.3422	83.9893	94.7108	89.3500	-117.2945	-103.9424	-110.6184
5	1	-0.5714	0.4422	-0.0646	-17.3896	-16.4484	-16.9190	2.0573	3.9225	2.9899	-14.3379	-12.4415	-13.3897	2.7104	3.7574	3.2339	-13.5802	-12.4349	-13.0076	3.5120	4.1076	3.8098	-12.5929	-11.9177	-12.2553
5	5	-42.8046	-26.5697	-34.6871	-77.5226	-63.8747	-70.6986	-20.3261	-0.9044	-10.6153	-42.5399	-26.2143	-34.3771	-7.1120	3.7129	-1.6996	-26.9667	-15.6419	-21.3043	-1.9042	5.9926	2.0442	-20.5769	-13.1147	-16.8458
5	13	26.5193	43.5142	35.0168	-374.7023	-357.6133	-366.1578	73.3284	98.3051	85.8167	-327.1026	-301.4440	-314.2733	78.5713	96.7640	87.6676	-321.3247	-303.0903	-312.2075	90.2498	103.5832	96.9165	-312.2444	-298.3966	-305.3205
6	1	0.4177	2.0879	1.2528	-32.6179	-30.9132	-31.7655	2.2868	3.2507	2.7688	-30.3682	-29.4527	-29.9105	2.9717	3.8212	3.3965	-29.3645	-28.5300	-28.9473	3.3282	4.0124	3.6703	-28.6510	-28.0461	-28.3485
6	5	-28.8235	-13.4456	-21.1345	-85.1657	-73.9455	-79.5556	-10.9369	0.1442	-5.3964	-53.6377	-43.5562	-48.5969	-3.8877	3.1056	-0.3910	-44.1082	-34.0584	-39.0833	-1.5163	3.7806	1.1322	-36.9507	-28.9514	-32.9511
6	13	35.1150	71.1977	53.1563	-766.2868	-729.4105	-747.8486	68.8481	89.3395	79.0938	-732.6240	-712.8486	-722.7363	80.6962	100.1471	90.4217	-721.3807	-701.2627	-711.3217	85.3869	99.7457	92.5663	-714.3169	-698.8602	-706.5885
7	1	-0.2475	1.3865	0.5695	-8.5214	-7.0814	-7.8014	2.3842	3.7960	3.0901	-5.9758	-4.7372	-5.3565	2.5421	3.4675	3.0048	-5.5390	-4.6490	-5.0940	3.4265	4.2228	3.8246	-4.5931	-3.8340	-4.2135
7	5	-39.9478	-17.4190	-28.6834	-65.8667	-47.9163	-56.8915	-10.0171	10.2036	0.0932	-26.2928	-8.0700	-17.1814	-6.4258	7.3786	0.4764	-20.8397	-7.8292	-14.3345	-4.5180	3.6836	-0.4172	-13.1399	-3.5060	-8.3229
7	13	35.8497	51.4400	43.6449	-154.3018	-130.9936	-142.6477	73.1229	89.9407	81.5318	-125.9857	-112.2166	-119.1011	81.3527	91.4047	86.3787	-117.9076	-107.3456	-112.6266	92.3451	103.3039	97.8245	-105.5353	-96.3170	-100.9262
8	1	0.4797	2.0608	1.2702	-16.0051	-13.7551	-14.8801	2.3811	3.3378	2.8595	-13.6066	-12.5760	-13.0913	3.3609	3.9542	3.6575	-12.7405	-12.0196	-12.3801	3.5463	4.1656	3.8559	-12.4833	-11.7374	-12.1103
8	5	-28.8978	-13.4031	-21.1505	-62.9985	-54.0583	-58.5284	-9.1062	2.4028	-3.3517	-35.2485	-22.1973	-28.7229	-1.7622	5.9724	2.1051	-23.1745	-14.3297	-18.7521	-0.7581	5.7365	2.4892	-18.3736	-9.9264	-14.1500
8	13	38.1582	68.1751	53.1667	-346.0400	-296.0913	-321.0657	70.5358	86.6270	78.5814	-321.1802	-308.0113	-314.5958	86.2653	98.2796	92.2725	-311.5522	-297.5992	-304.5757	89.3171	100.8629	95.0900	-309.4447	-295.3648	-302.4048
9	1	1.0305	2.4566	1.7436	-30.6660	-27.8716	-29.2688	2.6370	3.4918	3.0644	-29.6907	-28.2126	-28.9516	3.1842	3.7688	3.4765	-28.8536	-27.9603	-28.4069	3.4147	4.0220	3.7183	-28.6389	-27.9468	-28.2928
9	5	-17.2892	-2.7818	-10.0355	-76.4187	-53.8207	-65.1197	-4.5413	2.6575	-0.9419	-48.5234	-39.3697	-43.9465	-1.5168	3.8395	1.1613	-37.8862	-32.2425	-35.0644	1.5614	5.6137	3.5875	-33.3959	-28.8946	-31.1453
9	13	42.3520	67.2755	54.8137	-724.8830	-668.9301	-696.9065	73.3212	93.0729	83.1970	-720.6805	-686.5019	-703.5912	81.7536	94.4638	88.1087	-711.1564	-692.6430	-701.8997	87.3795	99.0791	93.2293	-710.6547	-697.2908	-703.9727

4.3 Paired-t Test and Sign Test

Paired-t test and sign test analyses were conducted on the results. All such tabulations are provided in Appendix A. Both tests determine whether or no a bootstrap method, as applied to the pseudo-experiment, improved the EV error component estiamte.

In the paired-t test, three outcomes are possible for each particular design, distributional set and replication level; using $\alpha = 0.05$.

1. If the CI contains zero then the EV and the bootstrap method used for comparison are considered the same; the EV is not improved.
2. If the CI contains negative values, then the bootstrap method has the best accuracy and should be used to perform the split-plot analysis, the method improves the EV.
3. If the CI contains positive values then EV estimate has the best accuracy and should be used to perform the split-plot analysis; the method does not improve the EV estimates.

The sign test is the non-parametric counter to the paired-t test but can yield additional insight in some cases. In the sign test, three outcomes are possible for each particular design, distributional set and replication level, using $\alpha = 0.12$.

1. If the p-value is between 0.06 and 0.94 then the EV and the bootstrap method used for comparison are considered the same.
2. If the p-value is less than 0.06 then the bootstrap method is the method that has the best accuracy and should be used to perform the split-plot analysis.
3. If the p-value is greater than 0.94 then EV is the method that has the best accuracy and should be used to perform the split-plot analysis.

In general, the re-sampling methods examined are not providing improved error component estimates. Additional inferences are made for two bootstrap methods,

RM2 and RM3. The results for both tests indicate that for a subset of designs and distributions the whole-plot error may in fact be estimated more accurately by RM2 and RM3 than just by EV. The subplot error estimates is still not as EV in these two methods. Further investigation is needed to confirm these findings and build upon the re-sampling methods presented.

V. Conclusions

5.1 Summary

Five bootstrap methods are defined and empirically examined to determine whether bootstrap methods can be used to improve the error component estimation in split-plot experiments. For the most part, the assessment of bootstrap as a viable methodology for improving the error estimation in split-plot designs is inconclusive. Of the five methods, none really provided consistent improvement over the analysis of just the experimental data. However, two methods (RM2 and RM3) did show promise in providing avenues to further research and for obtaining more accurate and precise estimates (At least for a subset of the conditions analyzed and reported on).

It is hoped that some of the details of this research can be useful to help drive theory behind the use of bootstrap methodology. That work can then in turn provide more detail in improving the accuracy and precision of the whole-plot error estimate.

Although the findings in this research were inconclusive, further investigation on additional re-sampling methods is warranted. A follow-on directly related can involve determining whether a bias correction could be applied to the whole-plot and subplot estimates found to improve accuracy. In addition, research on whether the whole-plot and subplot distributions estimation from the experimental data could be investigated.

The full realm of bootstrap methods have not been utilized and the use of any of the other methods discussed in the literature review may provide benefits to examining of split-plot analysis via bootstrap methodology. Future avenues of research include residual re-sampling methods that clarify v.s. obscure the error components. Observational sampling methods focused on purely increasing experimental size might show promise. Empirically looking at more varied distributional forms of the error components may yield insight into when re-sampling may be beneficial, a cursory assessment has been done, but not included. Finally, methods such as balanced bootstrap should be explored.

The following 270 tables summarize the EV estimates versus each re-sampling estimate for all designs for 3 distributions for the paired-t test and the sign test.

Table A.1: Paired-t Comparison - EV vs. RM1 - Design 1, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-8.8342	2.0807	Same
	3	-4.4024	0.1811	Same
	4	-7.4922	0.9897	Same
	5	1.5407	3.5140	EV
	6	-0.5681	2.7025	Same
	7	-0.0650	2.9114	Same
	8	-0.2798	3.8445	Same
	9	-1.0144	2.6441	Same
	10	1.6530	3.5701	EV
	15	-0.7366	1.9497	Same
	20	-0.1002	2.9043	Same
SP	2	-0.9373	1.0767	Same
	3	-0.4732	1.6936	Same
	4	-0.2483	2.1169	Same
	5	-0.4333	1.0018	Same
	6	0.1952	1.6511	EV
	7	0.3266	1.7970	EV
	8	0.0808	1.8473	EV
	9	0.2689	1.9850	EV
	10	0.9238	2.1534	EV
	15	1.5797	3.5052	EV
	20	2.2127	3.4809	EV

Table A.2: Sign Test Comparison - EV vs. RM1 - Design 1, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	11	0.252	Same
	3	12	0.132	Same
	4	12	0.132	Same
	5	3	0.999	EV
	6	8	0.748	Same
	7	5	0.979	EV
	8	5	0.979	EV
	9	7	0.868	Same
	10	1	1.000	EV
	15	9	0.588	Same
	20	2	1.000	EV
SP	2	12	0.132	Same
	3	8	0.748	Same
	4	9	0.588	Same
	5	10	0.412	Same
	6	4	0.994	EV
	7	5	0.979	EV
	8	6	0.942	EV
	9	4	0.994	EV
	10	3	0.999	EV
	15	2	1.000	EV
	20	0	1.000	EV

Table A.3: Paired-t Comparison - EV vs. RM1 - Design 1, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-78.0829	6.0980	Same
	3	-54.2687	-3.1332	RM1
	4	-77.2033	7.5955	Same
	5	-46.3246	9.3232	Same
	6	-44.6623	3.7665	Same
	7	-21.6526	0.8252	Same
	8	-36.2408	-2.4588	RM1
	9	-34.5185	2.1947	Same
	10	-17.5146	7.9761	Same
	15	-23.9004	-3.5784	RM1
	20	-20.0567	-0.9112	RM1
SP	2	-10.1069	20.5431	Same
	3	-23.9936	12.7778	Same
	4	-3.9523	18.3544	Same
	5	-12.6780	12.3256	Same
	6	-19.1234	6.9344	Same
	7	-14.1196	7.9666	Same
	8	-22.9675	0.2242	Same
	9	-22.3388	3.0970	Same
	10	-19.6819	1.1685	Same
	15	-13.8206	1.3918	Same
	20	-7.4064	2.9726	Same

Table A.4: Sign Test Comparison - EV vs. RM1 - Design 1, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	12	0.132	Same
	3	14	0.021	RM1
	4	13	0.058	RM1
	5	9	0.588	Same
	6	12	0.132	Same
	7	11	0.252	Same
	8	14	0.021	RM1
	9	14	0.021	RM1
	10	9	0.588	Same
	15	14	0.021	RM1
	20	14	0.021	RM1
SP	2	13	0.058	RM1
	3	10	0.412	Same
	4	5	0.979	EV
	5	7	0.868	Same
	6	11	0.252	Same
	7	10	0.412	Same
	8	12	0.132	Same
	9	13	0.058	RM1
	10	12	0.132	Same
	15	13	0.058	RM1
	20	14	0.021	RM1

Table A.5: Paired-t Comparison - EV vs. RM1 - Design 1, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-158.7413	36.3813	Same
	3	-27.5665	47.6953	Same
	4	-72.0783	70.4079	Same
	5	27.8511	80.3792	EV
	6	25.9156	76.7124	EV
	7	26.0799	76.2766	EV
	8	26.3686	75.2757	EV
	9	27.3655	82.2873	EV
	10	58.3210	92.1506	EV
	15	32.4933	72.9339	EV
	20	37.2710	80.0102	EV
SP	2	22.4520	60.6529	EV
	3	38.7233	73.9373	EV
	4	57.9539	93.2862	EV
	5	50.7435	80.5249	EV
	6	63.2956	85.4249	EV
	7	62.1370	88.9958	EV
	8	64.9518	80.5432	EV
	9	63.0225	83.7333	EV
	10	77.3504	94.9473	EV
	15	75.7828	100.2551	EV
	20	81.7744	98.6445	EV

Table A.6: Sign Test Comparison - EV vs. RM1 - Design 1, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	11	0.252	Same
	3	5	0.979	EV
	4	7	0.868	Same
	5	3	0.999	EV
	6	5	0.979	EV
	7	3	0.999	EV
	8	3	0.999	EV
	9	1	1.000	EV
	10	1	1.000	EV
	15	1	1.000	EV
	20	1	1.000	EV
SP	2	1	1.000	EV
	3	1	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.7: Paired-t Comparison - EV vs. RM1 - Design 2, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-3.8606	3.7479	Same
	3	0.2760	7.1793	EV
	4	1.9667	7.5838	EV
	5	-1.9305	8.5048	Same
	6	-1.3530	7.7338	Same
	7	-5.6373	7.4765	Same
	8	-2.6413	8.2008	Same
	9	2.9966	7.7833	EV
	10	3.7499	9.3356	EV
	15	2.1916	8.5492	EV
	20	7.9482	10.4434	EV
SP	2	-0.1104	1.7137	Same
	3	-0.1088	1.4784	Same
	4	-0.3862	1.0521	Same
	5	0.5367	2.1623	EV
	6	1.3917	3.3118	EV
	7	2.1032	4.6861	EV
	8	1.4147	3.2052	EV
	9	1.5500	3.1441	EV
	10	2.0188	3.8985	EV
	15	2.5738	3.9558	EV
	20	2.6015	3.6262	EV

Table A.8: Sign Test Comparison - EV vs. RM1 - Design 2, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	8	0.748	Same
	3	6	0.942	EV
	4	4	0.994	EV
	5	2	1.000	EV
	6	5	0.979	EV
	7	5	0.979	EV
	8	4	0.994	EV
	9	2	1.000	EV
	10	2	1.000	EV
	15	3	0.999	EV
	20	0	1.000	EV
SP	2	5	0.979	EV
	3	7	0.868	Same
	4	10	0.412	Same
	5	2	1.000	EV
	6	1	1.000	EV
	7	0	1.000	EV
	8	2	1.000	EV
	9	2	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.9: Paired-t Comparison - EV vs. RM1 - Design 2, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-62.8732	6.0723	Same
	3	-28.6715	2.6729	Same
	4	-36.5984	7.2023	Same
	5	-41.7084	5.3474	Same
	6	-44.5043	1.0445	Same
	7	-52.2701	5.7355	Same
	8	-51.0940	7.6508	Same
	9	-56.0630	-3.6432	RM1
	10	-37.8854	2.2678	Same
	15	-46.2438	-11.2699	RM1
	20	-22.0602	-0.4605	RM1
SP	2	-13.2920	25.6278	Same
	3	-6.9552	19.6034	Same
	4	-9.8028	11.0947	Same
	5	-1.7304	10.3827	Same
	6	-3.0849	7.0125	Same
	7	-3.9434	8.5530	Same
	8	-2.2966	7.6407	Same
	9	-4.7614	6.2082	Same
	10	-0.1616	7.8970	Same
	15	-6.7221	0.6580	Same
	20	-3.8660	2.4674	Same

Table A.10: Sign Test Comparison - EV vs. RM1 - Design 2, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	13	0.058	RM1
	3	12	0.132	Same
	4	13	0.058	RM1
	5	12	0.132	Same
	6	14	0.021	RM1
	7	11	0.252	Same
	8	13	0.058	RM1
	9	15	0.006	RM1
	10	12	0.132	Same
	15	16	0.001	RM1
	20	14	0.021	RM1
SP	2	7	0.868	Same
	3	6	0.942	EV
	4	6	0.942	EV
	5	7	0.868	Same
	6	7	0.868	Same
	7	8	0.748	Same
	8	8	0.748	Same
	9	8	0.748	Same
	10	8	0.748	Same
	15	14	0.021	RM1
	20	9	0.588	Same

Table A.11: Paired-t Comparison - EV vs. RM1 - Design 2, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	28.3720	161.3363	EV
	3	37.1779	154.1622	EV
	4	130.8455	239.2571	EV
	5	66.8165	253.6095	EV
	6	93.0902	231.6880	EV
	7	-36.9115	184.4578	Same
	8	50.4437	250.0125	EV
	9	149.2429	228.9644	EV
	10	86.3108	224.7952	EV
	15	67.3694	234.9967	EV
	20	194.3079	259.1705	EV
SP	2	39.3037	76.8297	EV
	3	38.8984	83.2675	EV
	4	46.4965	68.4372	EV
	5	55.4120	85.9368	EV
	6	74.0704	104.5303	EV
	7	76.7342	119.8519	EV
	8	66.6120	101.9517	EV
	9	70.0533	93.4502	EV
	10	82.9221	109.6857	EV
	15	82.9550	114.4904	EV
	20	83.2270	101.3938	EV

Table A.12: Sign Test Comparison - EV vs. RM1 - Design 2, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	4	0.994	EV
	3	4	0.994	EV
	4	1	1.000	EV
	5	1	1.000	EV
	6	2	1.000	EV
	7	3	0.999	EV
	8	2	1.000	EV
	9	0	1.000	EV
	10	3	0.999	EV
	15	2	1.000	EV
	20	1	1.000	EV
SP	2	1	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.13: Paired-t Comparison - EV vs. RM1 - Design 3, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-5.7936	15.8533	Same
	3	-20.8663	13.9322	Same
	4	-0.4069	17.2108	Same
	5	8.8662	17.8247	EV
	6	7.6440	20.0519	EV
	7	9.6456	16.7340	EV
	8	0.3730	20.1470	EV
	9	11.1407	21.0128	EV
	10	12.9298	21.6889	EV
	15	18.2111	24.1379	EV
	20	18.4876	24.2455	EV
SP	2	-0.3293	1.1048	Same
	3	1.2164	3.8101	EV
	4	1.0496	3.0933	EV
	5	1.6056	3.7489	EV
	6	1.6595	3.5825	EV
	7	2.0814	3.4580	EV
	8	2.3858	3.7390	EV
	9	1.9949	3.2000	EV
	10	2.5204	3.5894	EV
	15	2.7014	3.3404	EV
	20	3.0211	3.8276	EV

Table A.14: Sign Test Comparison - EV vs. RM1 - Design 3, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	7	0.868	Same
	3	5	0.979	EV
	4	5	0.979	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	0	1.000	EV
	8	3	0.999	EV
	9	3	0.999	EV
	10	1	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	8	0.748	Same
	3	3	0.999	EV
	4	3	0.999	EV
	5	2	1.000	EV
	6	3	0.999	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	1	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.15: Paired-t Comparison - EV vs. RM1 - Design 3, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-32.5971	9.4569	Same
	3	-133.6153	38.3772	Same
	4	-132.2620	-8.4491	RM1
	5	-56.2038	-4.2898	RM1
	6	-59.0476	-9.1194	RM1
	7	-34.3141	1.6295	Same
	8	-76.3053	9.1443	Same
	9	-42.5850	0.5325	Same
	10	-28.2139	3.2022	Same
	15	-16.0828	9.8107	Same
	20	-7.5930	12.1286	Same
SP	2	-19.1751	14.2020	Same
	3	-8.5819	7.3200	Same
	4	-6.3732	6.0263	Same
	5	-5.9009	4.7819	Same
	6	-3.8835	4.5090	Same
	7	-3.9153	4.7086	Same
	8	-2.0038	5.0828	Same
	9	-2.6283	2.1747	Same
	10	-3.0437	1.9695	Same
	15	-2.4853	0.6723	Same
	20	-1.4833	1.1206	Same

Table A.16: Sign Test Comparison - EV vs. RM1 - Design 3, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	13	0.058	RM1
	3	11	0.252	Same
	4	14	0.021	RM1
	5	14	0.021	RM1
	6	14	0.021	RM1
	7	12	0.132	Same
	8	12	0.132	Same
	9	12	0.132	Same
	10	13	0.058	RM1
	15	9	0.588	Same
	20	10	0.412	Same
SP	2	9	0.588	Same
	3	11	0.252	Same
	4	8	0.748	Same
	5	11	0.252	Same
	6	8	0.748	Same
	7	10	0.412	Same
	8	7	0.868	Same
	9	11	0.252	Same
	10	10	0.412	Same
	15	12	0.132	Same
	20	8	0.748	Same

Table A.17: Paired-t Comparison - EV vs. RM1 - Design 3, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-87.6710	356.7842	Same
	3	-275.0753	351.5137	Same
	4	146.0026	435.1887	EV
	5	279.2346	452.3615	EV
	6	289.0762	537.8988	EV
	7	267.4167	426.3876	EV
	8	161.8037	492.9599	EV
	9	354.1633	540.0879	EV
	10	354.6211	546.4037	EV
	15	474.0345	583.7742	EV
	20	456.1043	584.8268	EV
SP	2	29.2776	61.6995	EV
	3	50.1966	112.0131	EV
	4	56.5085	98.3661	EV
	5	60.5597	98.9128	EV
	6	61.9036	99.1688	EV
	7	71.6404	105.1350	EV
	8	78.0409	101.2988	EV
	9	63.9981	82.0375	EV
	10	82.8152	105.5262	EV
	15	78.9075	98.0466	EV
	20	87.2194	106.6894	EV

Table A.18: Sign Test Comparison - EV vs. RM1 - Design 3, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	4	0.994	EV
	3	5	0.979	EV
	4	1	1.000	EV
	5	0	1.000	EV
	6	1	1.000	EV
	7	0	1.000	EV
	8	2	1.000	EV
	9	0	1.000	EV
	10	2	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	1	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.19: Paired-t Comparison - EV vs. RM1 - Design 4, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-1.8902	3.8411	Same
	3	1.9422	4.4819	EV
	4	0.6442	3.9975	EV
	5	2.5717	4.1028	EV
	6	0.8069	3.7602	EV
	7	1.1681	3.9602	EV
	8	0.3689	3.1300	EV
	9	1.8141	3.3224	EV
	10	0.7795	3.2539	EV
	15	0.9344	2.8562	EV
	20	2.2485	3.2912	EV
SP	2	-0.7619	1.4862	Same
	3	0.0384	1.2198	EV
	4	-0.3104	1.2395	Same
	5	0.6583	1.7690	EV
	6	1.1149	2.9840	EV
	7	1.4538	3.1342	EV
	8	1.9072	3.4303	EV
	9	1.5319	2.7675	EV
	10	2.1337	3.3468	EV
	15	2.5014	3.8857	EV
	20	2.3261	3.2725	EV

Table A.20: Sign Test Comparison - EV vs. RM1 - Design 4, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	2	1.000	EV
	4	3	0.999	EV
	5	0	1.000	EV
	6	3	0.999	EV
	7	4	0.994	EV
	8	5	0.979	EV
	9	1	1.000	EV
	10	1	1.000	EV
	15	2	1.000	EV
	20	0	1.000	EV
SP	2	11	0.252	Same
	3	4	0.994	EV
	4	10	0.412	Same
	5	3	0.999	EV
	6	2	1.000	EV
	7	2	1.000	EV
	8	0	1.000	EV
	9	1	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.21: Paired-t Comparison - EV vs. RM1 - Design 4, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-49.4876	9.7995	Same
	3	-22.0396	15.0805	Same
	4	-18.6964	9.7079	Same
	5	-10.8633	15.6413	Same
	6	-17.9479	5.8583	Same
	7	-14.2808	7.5419	Same
	8	-12.9445	7.0963	Same
	9	-22.3277	-1.5249	RM1
	10	-17.4752	0.3426	Same
	15	-13.3491	4.3764	Same
	20	-12.9894	1.2985	Same
SP	2	-13.6614	18.6622	Same
	3	-5.5112	17.4631	Same
	4	-27.0135	-0.5132	RM1
	5	-19.0869	3.7475	Same
	6	-10.4372	4.2628	Same
	7	-8.5164	4.7462	Same
	8	-12.0120	1.5584	Same
	9	-12.6074	0.1960	Same
	10	-8.3842	4.8940	Same
	15	-7.5532	3.6056	Same
	20	-4.6129	3.3420	Same

Table A.22: Sign Test Comparison - EV vs. RM1 - Design 4, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	11	0.252	Same
	3	10	0.412	Same
	4	12	0.132	Same
	5	9	0.588	Same
	6	9	0.588	Same
	7	12	0.132	Same
	8	10	0.412	Same
	9	16	0.001	RM1
	10	12	0.132	Same
	15	9	0.588	Same
	20	12	0.132	Same
SP	2	8	0.748	Same
	3	8	0.748	Same
	4	12	0.132	Same
	5	13	0.058	RM1
	6	13	0.058	RM1
	7	11	0.252	Same
	8	13	0.058	RM1
	9	13	0.058	RM1
	10	12	0.132	Same
	15	10	0.412	Same
	20	11	0.252	Same

Table A.23: Paired-t Comparison - EV vs. RM1 - Design 4, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-1.0001	82.7309	Same
	3	60.6522	98.6997	EV
	4	45.4315	88.1692	EV
	5	77.3931	103.7942	EV
	6	48.0899	89.1110	EV
	7	46.5397	87.2122	EV
	8	28.2222	76.8255	EV
	9	68.6009	93.9498	EV
	10	57.9381	92.5300	EV
	15	55.8916	87.4106	EV
	20	77.9797	94.6621	EV
SP	2	39.5829	77.8544	EV
	3	54.3798	74.1430	EV
	4	58.8273	77.2860	EV
	5	73.2792	85.0377	EV
	6	71.8853	96.6837	EV
	7	73.9256	99.3655	EV
	8	78.1415	103.4767	EV
	9	73.9360	92.9302	EV
	10	80.1240	95.8414	EV
	15	89.0388	105.5934	EV
	20	84.9686	97.0421	EV

Table A.24: Sign Test Comparison - EV vs. RM1 - Design 4, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	4	0.994	EV
	3	0	1.000	EV
	4	2	1.000	EV
	5	0	1.000	EV
	6	1	1.000	EV
	7	2	1.000	EV
	8	2	1.000	EV
	9	0	1.000	EV
	10	1	1.000	EV
	15	1	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.25: Paired-t Comparison - EV vs. RM1 - Design 5, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	5.0799	10.5309	EV
	3	3.5464	9.3207	EV
	4	2.1183	9.5211	EV
	5	4.4399	10.9192	EV
	6	5.1331	9.5643	EV
	7	7.1747	11.3220	EV
	8	7.1015	9.5174	EV
	9	4.7830	9.9155	EV
	10	7.9405	10.6094	EV
	15	6.5481	10.4666	EV
	20	8.6739	10.6353	EV
SP	2	-0.5761	0.7271	Same
	3	0.9887	2.5585	EV
	4	1.2845	3.2084	EV
	5	2.1061	4.0220	EV
	6	2.2640	3.5218	EV
	7	2.3901	3.4037	EV
	8	2.4485	3.2886	EV
	9	2.7175	4.0388	EV
	10	2.6052	3.5985	EV
	15	2.8195	3.9303	EV
	20	3.3847	3.9999	EV

Table A.26: Sign Test Comparison - EV vs. RM1 - Design 5, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	2	1.000	EV
	3	1	1.000	EV
	4	2	1.000	EV
	5	1	1.000	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	0	1.000	EV
	9	2	1.000	EV
	10	0	1.000	EV
	15	1	1.000	EV
	20	0	1.000	EV
SP	2	9	0.588	Same
	3	3	0.999	EV
	4	3	0.999	EV
	5	1	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.27: Paired-t Comparison - EV vs. RM1 - Design 5, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-24.5011	29.5496	Same
	3	-31.4625	12.1798	Same
	4	-31.3644	11.4250	Same
	5	-39.2470	2.0997	Same
	6	-22.6134	2.8622	Same
	7	-13.6360	10.2683	Same
	8	-30.6054	-0.1084	RM1
	9	-14.6712	8.4451	Same
	10	-16.4031	2.4355	Same
	15	-22.7287	0.5187	Same
	20	-9.0016	5.4349	Same
SP	2	-33.6794	4.3878	Same
	3	-12.5619	5.5766	Same
	4	-8.3532	2.8016	Same
	5	-0.1029	9.4176	Same
	6	-7.5926	2.4755	Same
	7	-5.8462	2.8176	Same
	8	-5.1896	2.3133	Same
	9	-3.4539	3.0797	Same
	10	-4.3759	2.4540	Same
	15	-2.6248	2.4512	Same
	20	-2.3774	2.5589	Same

Table A.28: Sign Test Comparison - EV vs. RM1 - Design 5, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	8	0.748	Same
	3	9	0.588	Same
	4	10	0.412	Same
	5	12	0.132	Same
	6	14	0.021	RM1
	7	9	0.588	Same
	8	14	0.021	RM1
	9	9	0.588	Same
	10	11	0.252	Same
	15	12	0.132	Same
	20	10	0.412	Same
SP	2	11	0.252	Same
	3	9	0.588	Same
	4	11	0.252	Same
	5	6	0.942	EV
	6	12	0.132	Same
	7	11	0.252	Same
	8	11	0.252	Same
	9	10	0.412	Same
	10	11	0.252	Same
	15	8	0.748	Same
	20	11	0.252	Same

Table A.29: Paired-t Comparison - EV vs. RM1 - Design 5, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	165.1018	260.4982	EV
	3	165.3232	255.3646	EV
	4	102.9414	253.8565	EV
	5	167.4288	270.5802	EV
	6	197.3367	259.3694	EV
	7	200.2686	271.1835	EV
	8	229.6302	261.5236	EV
	9	154.6970	268.2379	EV
	10	201.7273	256.0610	EV
	15	191.9997	252.0770	EV
	20	229.9861	270.1536	EV
SP	2	37.2217	60.8457	EV
	3	63.7603	89.9941	EV
	4	64.8278	102.7821	EV
	5	80.7829	108.2935	EV
	6	77.9199	100.9863	EV
	7	79.6295	99.3812	EV
	8	79.9565	93.6243	EV
	9	89.6934	112.4910	EV
	10	82.5152	100.9435	EV
	15	85.9804	105.3781	EV
	20	91.5195	103.8624	EV

Table A.30: Sign Test Comparison - EV vs. RM1 - Design 5, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	0	1.000	EV
	3	1	1.000	EV
	4	2	1.000	EV
	5	1	1.000	EV
	6	0	1.000	EV
	7	1	1.000	EV
	8	0	1.000	EV
	9	1	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.31: Paired-t Comparison - EV vs. RM1 - Design 6, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	1.0101	19.3626	EV
	3	11.5598	19.9771	EV
	4	8.9737	23.6369	EV
	5	17.1540	23.6442	EV
	6	15.0908	21.6323	EV
	7	15.2584	21.9408	EV
	8	14.6152	21.7406	EV
	9	18.3538	24.5528	EV
	10	19.5002	24.4473	EV
	15	20.5216	24.0576	EV
	20	21.6079	24.1107	EV
SP	2	0.6535	3.3834	EV
	3	1.5315	3.1716	EV
	4	2.3856	3.6971	EV
	5	2.4663	3.5152	EV
	6	2.6681	3.9135	EV
	7	2.7972	4.0821	EV
	8	3.3411	4.4152	EV
	9	3.1474	4.1590	EV
	10	3.0239	3.8456	EV
	15	3.1405	4.1295	EV
	20	3.2764	3.9626	EV

Table A.32: Sign Test Comparison - EV vs. RM1 - Design 6, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	0	1.000	EV
	4	2	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	1	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	5	0.979	EV
	3	2	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.33: Paired-t Comparison - EV vs. RM1 - Design 6, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-77.1659	13.3165	Same
	3	-25.2448	9.2640	Same
	4	-59.5742	23.6315	Same
	5	-17.5053	11.0088	Same
	6	-19.1378	2.9685	Same
	7	-5.0875	21.0917	Same
	8	-24.9870	8.0207	Same
	9	-7.9759	17.9785	Same
	10	0.5501	16.7400	EV
	15	-8.4486	14.2460	Same
	20	5.8331	19.1232	EV
SP	2	-37.4138	-18.8144	RM1
	3	-17.3272	-6.6950	RM1
	4	-9.3720	-1.1754	RM1
	5	-6.8567	0.0798	Same
	6	-5.5842	-0.0521	RM1
	7	-4.1609	0.7430	Same
	8	-1.9038	0.9134	Same
	9	-3.1322	0.2613	Same
	10	-2.3500	0.7580	Same
	15	-1.1943	1.8697	Same
	20	-0.7013	1.6906	Same

Table A.34: Sign Test Comparison - EV vs. RM1 - Design 6, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	15	0.006	RM1
	3	10	0.412	Same
	4	8	0.748	Same
	5	10	0.412	Same
	6	12	0.132	Same
	7	3	0.999	EV
	8	10	0.412	Same
	9	8	0.748	Same
	10	6	0.942	EV
	15	7	0.868	Same
	20	3	0.999	EV
SP	2	17	0.000	RM1
	3	18	0.000	RM1
	4	14	0.021	RM1
	5	12	0.132	Same
	6	14	0.021	RM1
	7	13	0.058	RM1
	8	11	0.252	Same
	9	12	0.132	Same
	10	10	0.412	Same
	15	14	0.021	RM1
	20	10	0.412	Same

Table A.35: Paired-t Comparison - EV vs. RM1 - Design 6, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	96.1452	440.8126	EV
	3	353.2957	557.8097	EV
	4	336.9769	587.6258	EV
	5	417.2118	587.6477	EV
	6	416.0987	572.9633	EV
	7	440.6333	578.8209	EV
	8	433.3207	563.9421	EV
	9	489.0330	625.5607	EV
	10	498.7949	623.0076	EV
	15	548.0899	624.2569	EV
	20	539.6986	606.8424	EV
SP	2	54.9211	111.8214	EV
	3	62.6696	98.8585	EV
	4	75.6018	103.2683	EV
	5	78.2984	101.5574	EV
	6	77.4825	104.6064	EV
	7	83.3010	110.8233	EV
	8	90.2321	112.3993	EV
	9	88.6938	111.1728	EV
	10	85.7600	106.4786	EV
	15	86.0785	105.4459	EV
	20	87.7091	102.9253	EV

Table A.36: Sign Test Comparison - EV vs. RM1 - Design 6, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	0	1.000	EV
	4	1	1.000	EV
	5	1	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.37: Paired-t Comparison - EV vs. RM1 - Design 7, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	0.4968	4.8703	EV
	3	0.9012	3.9961	EV
	4	0.3692	3.8563	EV
	5	0.3594	4.3069	EV
	6	1.3027	4.2765	EV
	7	1.3833	3.4839	EV
	8	0.9165	3.2804	EV
	9	1.0844	3.8135	EV
	10	2.6334	3.6946	EV
	15	2.4073	3.6118	EV
	20	2.3005	3.4944	EV
SP	2	-0.3941	1.5817	Same
	3	1.2505	2.6949	EV
	4	1.7343	3.4409	EV
	5	2.2288	3.5822	EV
	6	2.4133	3.7656	EV
	7	2.1633	3.6573	EV
	8	2.5869	3.7836	EV
	9	2.6442	3.8565	EV
	10	2.2680	3.2232	EV
	15	3.2438	3.8537	EV
	20	3.1290	3.8709	EV

Table A.38: Sign Test Comparison - EV vs. RM1 - Design 7, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	4	0.994	EV
	4	5	0.979	EV
	5	3	0.999	EV
	6	3	0.999	EV
	7	2	1.000	EV
	8	4	0.994	EV
	9	3	0.999	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	1	1.000	EV
SP	2	9	0.588	Same
	3	2	1.000	EV
	4	2	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.39: Paired-t Comparison - EV vs. RM1 - Design 7, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-20.8259	16.0890	Same
	3	-23.0013	3.3087	Same
	4	-10.8819	15.0503	Same
	5	-25.1438	-0.7122	RM1
	6	-25.2313	1.0435	Same
	7	-18.6082	3.9297	Same
	8	-13.1623	4.5837	Same
	9	-16.5731	1.1186	Same
	10	-12.7552	1.9532	Same
	15	-13.0023	-0.1441	RM1
	20	-4.0300	5.0599	Same
SP	2	-48.1365	-2.1321	RM1
	3	-23.0479	-1.8255	RM1
	4	-13.9595	2.3414	Same
	5	-11.5743	3.6591	Same
	6	-6.7400	4.9525	Same
	7	-10.8028	4.4521	Same
	8	-9.7476	0.2330	Same
	9	-8.4811	0.1696	Same
	10	-6.5630	3.0083	Same
	15	-5.8586	1.5076	Same
	20	-7.2135	-0.1634	RM1

Table A.40: Sign Test Comparison - EV vs. RM1 - Design 7, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	7	0.868	Same
	3	11	0.252	Same
	4	6	0.942	EV
	5	13	0.058	RM1
	6	11	0.252	Same
	7	11	0.252	Same
	8	11	0.252	Same
	9	13	0.058	RM1
	10	12	0.132	Same
	15	14	0.021	RM1
	20	8	0.748	Same
SP	2	10	0.412	Same
	3	14	0.021	RM1
	4	13	0.058	RM1
	5	12	0.132	Same
	6	9	0.588	Same
	7	10	0.412	Same
	8	15	0.006	RM1
	9	16	0.001	RM1
	10	11	0.252	Same
	15	13	0.058	RM1
	20	12	0.132	Same

Table A.41: Paired-t Comparison - EV vs. RM1 - Design 7, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	44.7571	103.0558	EV
	3	45.4297	106.3880	EV
	4	22.4732	86.3414	EV
	5	57.5184	104.0195	EV
	6	60.7105	100.8043	EV
	7	64.3263	100.5166	EV
	8	49.2408	87.2549	EV
	9	71.9706	100.2067	EV
	10	83.3293	100.1966	EV
	15	75.3782	90.3044	EV
	20	72.1597	90.8698	EV
SP	2	61.2838	88.0397	EV
	3	72.5239	95.3174	EV
	4	74.5674	103.9180	EV
	5	82.1042	98.7225	EV
	6	84.2526	100.7955	EV
	7	86.2065	103.4308	EV
	8	91.1053	108.8517	EV
	9	86.1929	100.2587	EV
	10	86.2098	96.5400	EV
	15	92.3398	103.0861	EV
	20	95.4211	105.8765	EV

Table A.42: Sign Test Comparison - EV vs. RM1 - Design 7, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	2	1.000	EV
	3	1	1.000	EV
	4	4	0.994	EV
	5	1	1.000	EV
	6	1	1.000	EV
	7	2	1.000	EV
	8	2	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.43: Paired-t Comparison - EV vs. RM1 - Design 8, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-2.9173	8.8850	Same
	3	5.3939	11.1134	EV
	4	7.5351	10.3331	EV
	5	7.7910	10.4196	EV
	6	8.9593	11.0221	EV
	7	8.3127	10.7617	EV
	8	8.8500	10.7770	EV
	9	7.9910	10.2424	EV
	10	8.7816	10.8509	EV
	15	8.9558	11.0215	EV
	20	9.1455	10.4515	EV
SP	2	0.9525	3.6944	EV
	3	2.1089	3.5342	EV
	4	2.5304	3.3971	EV
	5	2.7927	3.6925	EV
	6	2.6451	3.2886	EV
	7	3.2487	3.9640	EV
	8	3.0700	3.8189	EV
	9	3.2433	4.0580	EV
	10	3.4613	4.0353	EV
	15	3.1901	3.8840	EV
	20	3.4580	4.0349	EV

Table A.44: Sign Test Comparison - EV vs. RM1 - Design 8, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	5	0.979	EV
	3	2	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	2	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.45: Paired-t Comparison - EV vs. RM1 - Design 8, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-60.9008	3.5631	Same
	3	-27.7200	2.4009	Same
	4	-32.2251	6.8098	Same
	5	-21.4603	1.7577	Same
	6	-18.1123	1.3111	Same
	7	-12.6331	4.7577	Same
	8	-22.9813	1.9556	Same
	9	-12.7833	3.8079	Same
	10	-7.7839	6.6123	Same
	15	-12.0767	-0.0462	RM1
	20	-6.1711	1.5744	Same
SP	2	-41.2444	-21.9801	RM1
	3	-19.2027	-5.6900	RM1
	4	-12.0543	-1.7573	RM1
	5	-7.3535	-0.5926	RM1
	6	-8.5520	-1.7893	RM1
	7	-4.7568	0.3948	Same
	8	-4.5685	1.6509	Same
	9	-4.0676	1.0689	Same
	10	-3.4633	1.3804	Same
	15	-2.2364	1.8524	Same
	20	-1.8198	2.4285	Same

Table A.46: Sign Test Comparison - EV vs. RM1 - Design 8, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	11	0.252	Same
	3	12	0.132	Same
	4	9	0.588	Same
	5	12	0.132	Same
	6	13	0.058	RM1
	7	10	0.412	Same
	8	10	0.412	Same
	9	12	0.132	Same
	10	7	0.868	Same
	15	12	0.132	Same
	20	12	0.132	Same
SP	2	18	0.000	RM1
	3	14	0.021	RM1
	4	15	0.006	RM1
	5	13	0.058	RM1
	6	15	0.006	RM1
	7	10	0.412	Same
	8	12	0.132	Same
	9	11	0.252	Same
	10	13	0.058	RM1
	15	11	0.252	Same
	20	7	0.868	Same

Table A.47: Paired-t Comparison - EV vs. RM1 - Design 8, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	10.3149	243.9461	EV
	3	178.9537	265.1991	EV
	4	235.2446	271.8359	EV
	5	232.8913	281.0747	EV
	6	243.0526	283.0975	EV
	7	225.9082	261.6161	EV
	8	247.9951	289.2988	EV
	9	225.3344	263.6771	EV
	10	229.9458	274.7659	EV
	15	249.3842	274.6658	EV
	20	241.1294	278.1106	EV
SP	2	69.9775	121.4506	EV
	3	78.0231	103.2014	EV
	4	81.3981	97.4178	EV
	5	84.4526	101.2106	EV
	6	83.5851	94.7544	EV
	7	93.2235	107.3547	EV
	8	85.8197	99.9356	EV
	9	92.5599	107.0972	EV
	10	92.7592	105.8365	EV
	15	86.4840	98.0285	EV
	20	92.3909	104.6055	EV

Table A.48: Sign Test Comparison - EV vs. RM1 - Design 8, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	1	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.49: Paired-t Comparison - EV vs. RM1 - Design 9, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	3.5276	23.6171	EV
	3	12.9496	22.1298	EV
	4	11.4646	22.1252	EV
	5	18.4770	24.4268	EV
	6	18.7151	26.5007	EV
	7	18.5793	24.3576	EV
	8	20.1466	24.9670	EV
	9	21.4638	25.3268	EV
	10	22.8095	25.1930	EV
	15	24.0792	25.8927	EV
	20	23.6567	25.8408	EV
SP	2	2.2648	4.1822	EV
	3	2.5894	4.1482	EV
	4	3.4729	4.6785	EV
	5	3.1496	4.0599	EV
	6	3.2842	4.1991	EV
	7	3.4967	4.4009	EV
	8	3.5356	4.4433	EV
	9	3.4659	4.1775	EV
	10	3.4202	4.0645	EV
	15	3.5016	3.9714	EV
	20	3.5465	4.0785	EV

Table A.50: Sign Test Comparison - EV vs. RM1 - Design 9, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	2	1.000	EV
	3	2	1.000	EV
	4	2	1.000	EV
	5	0	1.000	EV
	6	1	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.51: Paired-t Comparison - EV vs. RM1 - Design 9, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-90.1698	6.8050	Same
	3	-28.6034	3.4821	Same
	4	-34.5663	8.1219	Same
	5	-6.0917	14.2942	Same
	6	-10.9916	14.9294	Same
	7	-2.2351	20.3502	Same
	8	-3.1566	12.0869	Same
	9	-2.9640	15.7622	Same
	10	7.2684	22.1825	EV
	15	8.3534	20.6776	EV
	20	8.4710	17.5320	EV
SP	2	-26.3536	-18.0912	RM1
	3	-13.9076	-7.3992	RM1
	4	-5.8982	-1.0875	RM1
	5	-4.5721	-1.3379	RM1
	6	-3.4923	0.2143	Same
	7	-1.0772	1.6183	Same
	8	-2.0441	1.7813	Same
	9	-0.5032	2.4286	Same
	10	-0.4467	2.2376	Same
	15	-0.9158	1.5690	Same
	20	1.0562	2.8083	EV

Table A.52: Sign Test Comparison - EV vs. RM1 - Design 9, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	12	0.132	Same
	3	14	0.021	RM1
	4	9	0.588	Same
	5	9	0.588	Same
	6	6	0.942	EV
	7	5	0.979	EV
	8	7	0.868	Same
	9	9	0.588	Same
	10	4	0.994	EV
	15	4	0.994	EV
	20	1	1.000	EV
SP	2	20	0.000	RM1
	3	18	0.000	RM1
	4	17	0.000	RM1
	5	18	0.000	RM1
	6	13	0.058	RM1
	7	10	0.412	Same
	8	10	0.412	Same
	9	8	0.748	Same
	10	9	0.588	Same
	15	11	0.252	Same
	20	4	0.994	EV

Table A.53: Paired-t Comparison - EV vs. RM1 - Design 9, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	209.2620	587.7752	EV
	3	387.7583	577.5206	EV
	4	385.9712	575.8384	EV
	5	471.8090	612.6003	EV
	6	511.4795	650.9958	EV
	7	501.9256	621.1686	EV
	8	525.3030	642.5102	EV
	9	564.8506	649.1569	EV
	10	569.9054	631.9745	EV
	15	592.8742	630.0694	EV
	20	601.5254	648.3985	EV
SP	2	83.3165	123.7239	EV
	3	78.6869	112.4624	EV
	4	92.3433	117.6745	EV
	5	88.0339	111.5045	EV
	6	86.9012	107.4731	EV
	7	91.8877	111.3834	EV
	8	94.4471	114.5082	EV
	9	90.7280	106.1298	EV
	10	89.7400	103.9293	EV
	15	92.7448	104.7567	EV
	20	90.9994	103.0016	EV

Table A.54: Sign Test Comparison - EV vs. RM1 - Design 9, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	1	1.000	EV
	3	1	1.000	EV
	4	1	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.55: Paired-t Comparison - EV vs. RM2 - Design 1, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-0.5713	0.0584	Same
	3	-1.1442	0.2399	Same
	4	-0.8415	0.4605	Same
	5	-0.9415	0.4995	Same
	6	-0.7014	0.8537	Same
	7	-1.1119	0.3618	Same
	8	-1.3572	0.2490	Same
	9	-1.5932	0.0209	Same
	10	-1.1546	0.0624	Same
	15	-1.4965	0.2545	Same
	20	-1.2898	0.3339	Same
SP	2	0.6394	1.6627	EV
	3	0.2422	1.7604	EV
	4	1.0740	1.9175	EV
	5	0.9091	1.8208	EV
	6	0.6862	1.6621	EV
	7	0.9077	1.8462	EV
	8	0.5409	1.5792	EV
	9	0.0833	1.7340	EV
	10	0.3994	1.6661	EV
	15	0.6405	1.7693	EV
	20	1.0028	1.7780	EV

Table A.56: Sign Test Comparison - EV vs. RM2 - Design 1, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	14	0.021	RM2
	3	14	0.021	RM2
	4	12	0.132	Same
	5	12	0.132	Same
	6	10	0.412	Same
	7	12	0.132	Same
	8	13	0.058	RM2
	9	14	0.021	RM2
	10	14	0.021	RM2
	15	13	0.058	RM2
	20	13	0.058	RM2
SP	2	1	1.000	EV
	3	3	0.999	EV
	4	2	1.000	EV
	5	3	0.999	EV
	6	1	1.000	EV
	7	1	1.000	EV
	8	3	0.999	EV
	9	3	0.999	EV
	10	2	1.000	EV
	15	3	0.999	EV
	20	2	1.000	EV

Table A.57: Paired-t Comparison - EV vs. RM2 - Design 1, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-12.1180	2.5998	Same
	3	-18.8563	17.0244	Same
	4	-13.1398	15.2133	Same
	5	-19.6939	9.5347	Same
	6	-16.0543	15.8088	Same
	7	-17.4348	14.5662	Same
	8	-14.6661	18.4920	Same
	9	-23.4590	18.7443	Same
	10	-18.3466	16.5002	Same
	15	-10.9290	20.0704	Same
	20	12.0651	42.5051	EV
SP	2	15.8679	41.4739	EV
	3	5.5633	43.8181	EV
	4	27.2419	48.0191	EV
	5	22.9600	45.5374	EV
	6	16.6445	41.1103	EV
	7	22.6497	46.0131	EV
	8	13.5624	39.4791	EV
	9	1.9928	43.0768	EV
	10	9.5054	41.5523	EV
	15	16.1813	44.1472	EV
	20	24.9259	44.2555	EV

Table A.58: Sign Test Comparison - EV vs. RM2 - Design 1, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	12	0.132	Same
	3	12	0.132	Same
	4	10	0.412	Same
	5	13	0.058	RM2
	6	10	0.412	Same
	7	11	0.252	Same
	8	8	0.748	Same
	9	12	0.132	Same
	10	10	0.412	Same
	15	10	0.412	Same
	20	6	0.942	EV
SP	2	1	1.000	EV
	3	3	0.999	EV
	4	2	1.000	EV
	5	3	0.999	EV
	6	1	1.000	EV
	7	1	1.000	EV
	8	3	0.999	EV
	9	3	0.999	EV
	10	2	1.000	EV
	15	3	0.999	EV
	20	1	1.000	EV

Table A.59: Paired-t Comparison - EV vs. RM2 - Design 1, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-0.5813	0.4004	Same
	3	-0.9251	0.7897	Same
	4	-0.4004	1.0410	Same
	5	-1.5136	-0.0872	RM2
	6	-1.0185	0.6685	Same
	7	-0.9216	0.8001	Same
	8	-1.3055	0.2668	Same
	9	-2.0774	-0.4929	RM2
	10	-0.8398	0.9993	Same
	15	-1.7074	-0.0153	RM2
	20	-1.9487	-0.2923	RM2
SP	2	0.6585	1.6659	EV
	3	0.2419	1.7782	EV
	4	1.0653	1.9116	EV
	5	0.8978	1.8200	EV
	6	0.6864	1.6578	EV
	7	0.9122	1.8476	EV
	8	0.5138	1.5635	EV
	9	0.0848	1.7354	EV
	10	0.3870	1.6747	EV
	15	0.6506	1.7716	EV
	20	1.0061	1.7747	EV

Table A.60: Sign Test Comparison - EV vs. RM2 - Design 1, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	14	0.021	RM2
	3	15	0.006	RM2
	4	9	0.588	Same
	5	15	0.006	RM2
	6	13	0.058	RM2
	7	11	0.252	Same
	8	13	0.058	RM2
	9	16	0.001	RM2
	10	10	0.412	Same
	15	14	0.021	RM2
	20	15	0.006	RM2
SP	2	1	1.000	EV
	3	3	0.999	EV
	4	2	1.000	EV
	5	2	1.000	EV
	6	1	1.000	EV
	7	1	1.000	EV
	8	3	0.999	EV
	9	3	0.999	EV
	10	2	1.000	EV
	15	3	0.999	EV
	20	1	1.000	EV

Table A.61: Paired-t Comparison - EV vs. RM2 - Design 2, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-1.3425	0.2963	Same
	3	-1.8712	-0.2237	RM2
	4	-2.0533	-0.2956	RM2
	5	-1.4261	0.6567	Same
	6	-0.6410	1.6299	Same
	7	-1.0451	1.2718	Same
	8	-1.4911	0.8965	Same
	9	-2.1454	0.2667	Same
	10	-1.1149	1.2861	Same
	15	-0.9146	1.5592	Same
	20	-0.9899	1.6044	Same
SP	2	0.0481	1.6364	EV
	3	0.1348	1.3583	EV
	4	0.2740	1.2096	EV
	5	0.7316	1.2387	EV
	6	0.3965	1.1884	EV
	7	0.7224	1.2205	EV
	8	0.6965	1.1615	EV
	9	0.2864	0.9270	EV
	10	0.6119	1.1204	EV
	15	0.4975	0.9761	EV
	20	0.5208	0.9497	EV

Table A.62: Sign Test Comparison - EV vs. RM2 - Design 2, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	15	0.006	RM2
	3	15	0.006	RM2
	4	15	0.006	RM2
	5	12	0.132	Same
	6	8	0.748	Same
	7	11	0.252	Same
	8	11	0.252	Same
	9	14	0.021	RM2
	10	10	0.412	Same
	15	9	0.588	Same
	20	9	0.588	Same
SP	2	4	0.994	EV
	3	3	0.999	EV
	4	3	0.999	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	2	1.000	EV
	9	3	0.999	EV
	10	2	1.000	EV
	15	2	1.000	EV
	20	2	1.000	EV

Table A.63: Paired-t Comparison - EV vs. RM2 - Design 2, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-26.8307	9.7778	Same
	3	-35.6989	-0.7951	RM2
	4	-28.8075	11.7473	Same
	5	-12.4041	35.6159	Same
	6	-4.9280	43.6083	Same
	7	8.4626	51.9790	EV
	8	-10.3340	42.8150	Same
	9	-15.0189	34.5258	Same
	10	-2.8228	45.9169	Same
	15	4.6458	48.3749	EV
	20	30.9298	60.8950	EV
SP	2	1.1664	40.8060	EV
	3	3.4258	33.7132	EV
	4	6.9498	30.1481	EV
	5	18.1453	30.8348	EV
	6	9.8841	29.8009	EV
	7	18.0232	30.3464	EV
	8	17.2524	29.0190	EV
	9	6.7786	23.0426	EV
	10	15.3976	27.9313	EV
	15	12.2646	24.3271	EV
	20	12.9564	23.7890	EV

Table A.64: Sign Test Comparison - EV vs. RM2 - Design 2, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	13	0.058	RM2
	3	15	0.006	RM2
	4	11	0.252	Same
	5	7	0.868	Same
	6	8	0.748	Same
	7	4	0.994	EV
	8	8	0.748	Same
	9	8	0.748	RM2
	10	6	0.942	EV
	15	4	0.994	EV
	20	3	0.999	EV
SP	2	4	0.994	EV
	3	3	0.999	EV
	4	3	0.999	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	2	1.000	EV
	9	3	0.999	EV
	10	2	1.000	EV
	15	2	1.000	EV
	20	2	1.000	EV

Table A.65: Paired-t Comparison - EV vs. RM2 - Design 2, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-1.2288	1.0005	Same
	3	-1.8029	-0.1222	RM2
	4	-2.4756	-0.3162	RM2
	5	-0.7211	1.2470	Same
	6	-1.1404	1.2677	Same
	7	-1.5814	0.9296	Same
	8	-1.5837	1.2519	Same
	9	-2.0158	0.5971	Same
	10	-0.8628	1.6390	Same
	15	-1.7687	0.9454	Same
	20	-1.5330	1.1649	Same
SP	2	0.0560	1.6363	EV
	3	0.1422	1.3554	EV
	4	0.2785	1.2236	EV
	5	0.7295	1.2417	EV
	6	0.3930	1.1943	EV
	7	0.7223	1.2168	EV
	8	0.6967	1.1618	EV
	9	0.2911	0.9351	EV
	10	0.6118	1.1137	EV
	15	0.4862	0.9728	EV
	20	0.5185	0.9532	EV

Table A.66: Sign Test Comparison - EV vs. RM2 - Design 2, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	14	0.021	RM2
	3	15	0.006	RM2
	4	16	0.001	RM2
	5	10	0.412	Same
	6	10	0.412	Same
	7	11	0.252	Same
	8	10	0.412	Same
	9	13	0.058	RM2
	10	9	0.588	Same
	15	11	0.252	Same
	20	11	0.252	Same
SP	2	4	0.994	EV
	3	3	0.999	EV
	4	3	0.999	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	2	1.000	EV
	9	3	0.999	EV
	10	2	1.000	EV
	15	2	1.000	EV
	20	2	1.000	EV

Table A.67: Paired-t Comparison - EV vs. RM2 - Design 3, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-1.4449	0.6612	Same
	3	-1.6715	0.4882	Same
	4	-0.9380	1.9341	Same
	5	-2.2204	0.6839	Same
	6	-1.2967	1.8468	Same
	7	-2.5513	0.3455	Same
	8	-0.5715	2.4315	Same
	9	-2.1666	0.7807	Same
	10	-1.8530	1.1333	Same
	15	-1.8414	1.0803	Same
	20	-1.3691	1.5870	Same
SP	2	-0.6828	0.7769	Same
	3	-0.1852	0.7167	Same
	4	-0.1774	0.6571	Same
	5	-0.2323	0.4766	Same
	6	-0.1575	0.5700	Same
	7	-0.0417	0.6184	Same
	8	-0.1737	0.4547	Same
	9	-0.1610	0.4787	Same
	10	-0.0532	0.4307	Same
	15	-0.0785	0.3784	Same
	20	0.0617	0.4062	EV

Table A.68: Sign Test Comparison - EV vs. RM2 - Design 3, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	13	0.058	RM2
	3	12	0.132	RM2
	4	9	0.588	Same
	5	13	0.058	RM2
	6	10	0.412	Same
	7	14	0.021	RM2
	8	8	0.748	Same
	9	12	0.132	Same
	10	11	0.252	Same
	15	11	0.252	Same
	20	9	0.588	Same
SP	2	7	0.868	Same
	3	8	0.748	Same
	4	6	0.942	EV
	5	9	0.588	Same
	6	8	0.748	Same
	7	6	0.942	EV
	8	8	0.748	Same
	9	8	0.748	Same
	10	7	0.868	Same
	15	7	0.868	Same
	20	6	0.942	EV

Table A.69: Paired-t Comparison - EV vs. RM2 - Design 3, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-45.1745	0.3150	Same
	3	-19.4101	34.3183	Same
	4	-19.1879	43.8650	Same
	5	3.2432	60.8828	EV
	6	11.4539	70.1417	EV
	7	9.0862	70.6987	EV
	8	24.3736	75.1638	EV
	9	5.1970	57.4637	EV
	10	2.3723	62.2594	EV
	15	39.3881	85.8203	EV
	20	48.1957	82.1777	EV
SP	2	-17.0711	19.4229	Same
	3	-4.6312	17.9179	Same
	4	-4.4355	16.4269	Same
	5	-5.8074	11.9151	Same
	6	-3.9373	14.2494	Same
	7	-1.0425	15.4591	Same
	8	-4.3433	11.3673	Same
	9	-4.0250	11.9675	Same
	10	-1.3301	10.7683	Same
	15	-1.9617	9.4589	Same
	20	1.5434	10.1562	EV

Table A.70: Sign Test Comparison - EV vs. RM2 - Design 3, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	15	0.006	RM2
	3	9	0.588	Same
	4	9	0.588	Same
	5	6	0.942	EV
	6	6	0.942	EV
	7	5	0.979	EV
	8	5	0.979	EV
	9	7	0.868	Same
	10	5	0.979	EV
	15	3	0.999	EV
	20	1	1.000	EV
SP	2	7	0.868	Same
	3	8	0.748	Same
	4	6	0.942	EV
	5	9	0.588	Same
	6	8	0.748	Same
	7	6	0.942	EV
	8	8	0.748	Same
	9	8	0.748	Same
	10	7	0.868	Same
	15	7	0.868	Same
	20	6	0.942	EV

Table A.71: Paired-t Comparison - EV vs. RM2 - Design 3, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-2.3176	1.2231	Same
	3	-2.2825	-0.4496	RM2
	4	-2.2698	1.3128	Same
	5	-2.8104	0.2108	Same
	6	-3.2818	0.1510	Same
	7	-2.3584	1.1384	Same
	8	-1.1160	2.5017	Same
	9	-2.9769	-0.3724	RM2
	10	-2.2728	0.8245	Same
	15	-3.0000	-0.2259	RM2
	20	-2.9336	-0.1599	RM2
SP	2	-0.6828	0.7769	Same
	3	-0.1852	0.7167	Same
	4	-0.1774	0.6571	Same
	5	-0.2323	0.4766	Same
	6	-0.1575	0.5700	Same
	7	-0.0417	0.6184	Same
	8	-0.1737	0.4547	Same
	9	-0.1610	0.4787	Same
	10	-0.0532	0.4307	Same
	15	-0.0785	0.3784	Same
	20	0.0617	0.4062	EV

Table A.72: Sign Test Comparison - EV vs. RM2 - Design 3, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	16	0.001	RM2
	3	18	0.000	RM2
	4	12	0.132	Same
	5	14	0.021	RM2
	6	14	0.021	RM2
	7	13	0.058	RM2
	8	8	0.748	Same
	9	15	0.006	RM2
	10	12	0.132	Same
	15	15	0.006	RM2
	20	15	0.006	RM2
SP	2	7	0.868	Same
	3	8	0.748	Same
	4	6	0.942	EV
	5	9	0.588	Same
	6	8	0.748	Same
	7	6	0.942	EV
	8	8	0.748	Same
	9	8	0.748	Same
	10	7	0.868	Same
	15	7	0.868	Same
	20	6	0.942	EV

Table A.73: Paired-t Comparison - EV vs. RM2 - Design 4, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-0.8724	0.2342	Same
	3	-0.8366	0.3668	Same
	4	-1.3603	0.0456	Same
	5	-0.8142	0.6735	Same
	6	-0.9848	0.5424	Same
	7	-0.4963	0.9907	Same
	8	-0.7731	0.8120	Same
	9	-1.3704	0.1258	Same
	10	-0.9126	0.8352	Same
	15	-0.6729	1.0074	Same
	20	-1.0359	0.7577	Same
SP	2	0.5026	1.7847	EV
	3	1.0719	1.8652	EV
	4	0.2611	1.4734	EV
	5	0.2227	1.7323	EV
	6	0.7119	1.8174	EV
	7	0.9091	1.8947	EV
	8	1.0994	1.9669	EV
	9	0.8417	1.7359	EV
	10	0.9365	1.8161	EV
	15	1.1615	1.8399	EV
	20	1.3300	1.7986	EV

Table A.74: Sign Test Comparison - EV vs. RM2 - Design 4, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	13	0.058	RM2
	3	12	0.132	Same
	4	14	0.021	RM2
	5	10	0.412	Same
	6	12	0.132	Same
	7	9	0.588	Same
	8	9	0.588	Same
	9	14	0.021	RM2
	10	11	0.252	Same
	15	10	0.412	Same
	20	11	0.252	Same
SP	2	2	1.000	EV
	3	1	1.000	EV
	4	5	0.979	EV
	5	3	0.999	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	2	1.000	EV
	9	2	1.000	EV
	10	2	1.000	EV
	15	2	1.000	EV
	20	0	1.000	EV

Table A.75: Paired-t Comparison - EV vs. RM2 - Design 4, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-21.5000	0.1493	Same
	3	-26.3814	-0.6414	RM2
	4	-14.2285	17.8802	Same
	5	-11.9716	25.5195	Same
	6	-12.7447	21.7528	Same
	7	-4.0847	29.6417	Same
	8	-0.2614	25.6110	Same
	9	-7.8575	24.8676	Same
	10	7.1937	40.2851	EV
	15	7.2258	38.3678	EV
	20	14.1829	41.0902	EV
SP	2	0.5026	1.7847	EV
	3	1.0719	1.8652	EV
	4	0.2611	1.4734	EV
	5	0.2227	1.7323	EV
	6	0.7119	1.8174	EV
	7	0.9091	1.8947	EV
	8	1.0994	1.9669	EV
	9	0.8417	1.7359	EV
	10	0.9365	1.8161	EV
	15	1.1615	1.8399	EV
	20	1.3300	1.7986	EV

Table A.76: Sign Test Comparison - EV vs. RM2 - Design 4, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	15	0.006	RM2
	3	15	0.006	RM2
	4	10	0.412	Same
	5	8	0.748	Same
	6	9	0.588	Same
	7	7	0.868	Same
	8	5	0.979	EV
	9	10	0.412	Same
	10	6	0.942	EV
	15	5	0.979	EV
	20	4	0.994	EV
SP	2	2	1.000	EV
	3	1	1.000	EV
	4	5	0.979	EV
	5	3	0.999	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	2	1.000	EV
	9	2	1.000	EV
	10	2	1.000	EV
	15	2	1.000	EV
	20	0	1.000	EV

Table A.77: Paired-t Comparison - EV vs. RM2 - Design 4, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-0.9095	0.2776	Same
	3	-0.7486	0.4895	Same
	4	-1.4296	0.0563	Same
	5	-1.1714	0.4296	Same
	6	-1.0383	0.6935	Same
	7	-0.9748	0.6496	Same
	8	-1.1347	0.5531	Same
	9	-1.4049	0.3389	Same
	10	-1.4608	0.3285	Same
	15	-1.3598	0.4810	Same
	20	-1.6044	0.1692	Same
SP	2	0.5001	1.7813	EV
	3	1.0674	1.8641	EV
	4	0.2636	1.4710	EV
	5	0.2015	1.7240	EV
	6	0.7161	1.8281	EV
	7	0.8916	1.8914	EV
	8	1.1038	1.9646	EV
	9	0.8488	1.7344	EV
	10	0.9398	1.8163	EV
	15	1.1666	1.8372	EV
	20	1.3268	1.7972	EV

Table A.78: Sign Test Comparison - EV vs. RM2 - Design 4, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	13	0.058	RM2
	3	11	0.252	Same
	4	14	0.021	RM2
	5	12	0.132	Same
	6	12	0.132	Same
	7	12	0.132	Same
	8	11	0.252	Same
	9	14	0.021	RM2
	10	13	0.058	RM2
	15	12	0.132	Same
	20	14	0.021	RM2
SP	2	3	0.999	EV
	3	1	1.000	EV
	4	5	0.979	EV
	5	3	0.999	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	2	1.000	EV
	9	2	1.000	EV
	10	2	1.000	EV
	15	2	1.000	EV
	20	0	1.000	EV

Table A.79: Paired-t Comparison - EV vs. RM2 - Design 5, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-1.6105	0.2615	Same
	3	-1.1661	1.1102	Same
	4	-1.2651	0.9873	Same
	5	-0.0200	2.0906	Same
	6	-2.0297	0.4053	Same
	7	-0.3996	1.8011	Same
	8	-1.4737	1.1114	Same
	9	-1.2716	1.1461	Same
	10	-0.6390	1.8002	Same
	15	-0.5493	1.8238	Same
	20	-0.0815	2.0992	Same
SP	2	-0.9891	0.8246	Same
	3	-0.2626	0.9292	Same
	4	0.2394	0.9817	EV
	5	0.3082	1.1287	EV
	6	-0.2745	0.7086	Same
	7	0.2368	0.8621	EV
	8	0.1376	0.8472	EV
	9	0.4196	0.9246	EV
	10	0.3601	0.9059	EV
	15	0.4202	0.8426	EV
	20	0.5347	0.8607	EV

Table A.80: Sign Test Comparison - EV vs. RM2 - Design 5, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	12	0.132	Same
	3	11	0.252	Same
	4	10	0.412	Same
	5	6	0.942	EV
	6	14	0.021	RM2
	7	9	0.588	Same
	8	11	0.252	Same
	9	10	0.412	Same
	10	8	0.748	Same
	15	8	0.748	Same
	20	6	0.942	EV
SP	2	9	0.588	Same
	3	5	0.979	EV
	4	3	0.999	EV
	5	4	0.994	EV
	6	8	0.748	Same
	7	5	0.979	EV
	8	4	0.994	EV
	9	4	0.994	EV
	10	5	0.979	EV
	15	2	1.000	EV
	20	0	1.000	EV

Table A.81: Paired-t Comparison - EV vs. RM2 - Design 5, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-24.1736	17.5170	Same
	3	-14.8764	34.7164	Same
	4	-1.7524	43.5398	Same
	5	1.5599	47.3601	EV
	6	17.5029	61.2001	EV
	7	31.6004	63.8264	EV
	8	20.5953	59.2564	EV
	9	21.5863	63.4977	EV
	10	32.6709	64.8253	EV
	15	42.4199	66.9122	EV
	20	46.8418	72.9529	EV
SP	2	-24.7275	20.6145	Same
	3	-6.5644	23.2307	Same
	4	5.9842	24.5433	EV
	5	7.7060	28.2187	EV
	6	-6.8624	17.7140	Same
	7	5.9200	21.5537	EV
	8	3.4407	21.1798	EV
	9	10.4898	23.1138	EV
	10	9.0032	22.6471	EV
	15	10.5061	21.0642	EV
	20	13.3672	21.5176	EV

Table A.82: Sign Test Comparison - EV vs. RM2 - Design 5, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	10	0.412	Same
	3	9	0.588	Same
	4	7	0.868	Same
	5	6	0.942	EV
	6	5	0.979	EV
	7	3	0.999	EV
	8	4	0.994	EV
	9	4	0.994	EV
	10	2	1.000	EV
	15	1	1.000	EV
	20	1	1.000	EV
SP	2	9	0.588	Same
	3	5	0.979	EV
	4	3	0.999	EV
	5	4	0.994	EV
	6	8	0.748	Same
	7	5	0.979	EV
	8	4	0.994	EV
	9	4	0.994	EV
	10	5	0.979	EV
	15	2	1.000	EV
	20	0	1.000	EV

Table A.83: Paired-t Comparison - EV vs. RM2 - Design 5, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-2.0460	-0.4462	Same
	3	-1.8581	0.8444	Same
	4	-1.5242	1.0447	Same
	5	-0.9038	1.5868	Same
	6	-1.9663	0.7933	Same
	7	-1.7709	0.8912	Same
	8	-1.9893	0.5591	Same
	9	-1.4443	1.0875	Same
	10	-1.5535	1.1076	Same
	15	-1.8250	1.0157	Same
	20	-2.0426	0.7579	Same
SP	2	-0.9974	0.8206	Same
	3	-0.2353	0.9439	Same
	4	0.2316	0.9787	EV
	5	0.2994	1.1250	EV
	6	-0.2569	0.7126	Same
	7	0.2452	0.8724	EV
	8	0.1433	0.8500	EV
	9	0.4163	0.9255	EV
	10	0.3694	0.9087	EV
	15	0.4273	0.8467	EV
	20	0.5318	0.8613	EV

Table A.84: Sign Test Comparison - EV vs. RM2 - Design 5, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	16	0.001	RM2
	3	12	0.132	Same
	4	11	0.252	Same
	5	9	0.588	Same
	6	12	0.132	Same
	7	11	0.252	Same
	8	12	0.132	Same
	9	11	0.252	Same
	10	11	0.252	Same
	15	12	0.132	Same
	20	12	0.132	Same
SP	2	9	0.588	Same
	3	5	0.979	EV
	4	3	0.999	EV
	5	4	0.994	EV
	6	8	0.748	Same
	7	5	0.979	EV
	8	4	0.994	EV
	9	4	0.994	EV
	10	5	0.979	EV
	15	2	1.000	EV
	20	0	1.000	EV

Table A.85: Paired-t Comparison - EV vs. RM2 - Design 6, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-1.3408	1.2069	Same
	3	-1.7470	1.0751	Same
	4	-1.3151	1.5259	Same
	5	-2.4984	0.0929	Same
	6	-1.7979	1.2672	Same
	7	-1.7221	1.3889	Same
	8	-0.2424	2.6939	Same
	9	0.6147	2.9854	EV
	10	-1.7699	1.0838	Same
	15	-1.2408	1.8812	Same
	20	-1.4846	1.7286	Same
SP	2	-1.8497	-0.7991	RM2
	3	-1.0684	-0.2721	RM2
	4	-0.6539	0.0747	Same
	5	-0.5307	0.1572	Same
	6	-0.3548	0.2892	Same
	7	-0.3097	0.2807	Same
	8	-0.3231	0.2227	Same
	9	-0.2361	0.3154	Same
	10	-0.1622	0.3030	Same
	15	0.0373	0.3607	EV
	20	0.1300	0.3951	EV

Table A.86: Sign Test Comparison - EV vs. RM2 - Design 6, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	9	0.588	Same
	3	11	0.252	Same
	4	10	0.412	Same
	5	14	0.021	RM2
	6	10	0.412	Same
	7	11	0.252	Same
	8	6	0.942	EV
	9	4	0.994	EV
	10	12	0.132	Same
	15	9	0.588	Same
	20	10	0.412	Same
SP	2	17	0.000	RM2
	3	16	0.001	RM2
	4	12	0.132	Same
	5	12	0.132	Same
	6	9	0.588	Same
	7	10	0.412	Same
	8	11	0.252	Same
	9	10	0.412	Same
	10	8	0.748	Same
	15	8	0.748	Same
	20	4	0.994	EV

Table A.87: Paired-t Comparison - EV vs. RM2 - Design 6, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-16.1120	40.9905	Same
	3	21.4499	70.2240	EV
	4	28.2283	71.5229	EV
	5	10.2172	62.4613	EV
	6	25.7528	74.9430	EV
	7	38.8198	76.6351	EV
	8	54.3135	89.0414	EV
	9	56.7945	81.7486	EV
	10	54.0547	82.4583	EV
	15	50.0930	77.9686	EV
	20	56.4722	83.7865	EV
SP	2	-46.2432	-19.9770	RM2
	3	-26.7110	-6.8031	RM2
	4	-16.3476	1.8668	Same
	5	-13.2682	3.9292	Same
	6	-8.8707	7.2301	Same
	7	-7.7415	7.0176	Same
	8	-8.0771	5.5670	Same
	9	-5.9024	7.8856	Same
	10	-4.0557	7.5743	Same
	15	0.9324	9.0171	EV
	20	3.2507	9.8785	EV

Table A.88: Sign Test Comparison - EV vs. RM2 - Design 6, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	8	0.748	Same
	3	5	0.979	EV
	4	3	0.999	EV
	5	5	0.979	EV
	6	4	0.994	EV
	7	2	1.000	EV
	8	1	1.000	EV
	9	1	1.000	EV
	10	1	1.000	EV
	15	1	1.000	EV
	20	1	1.000	EV
SP	2	17	0.000	RM2
	3	16	0.001	RM2
	4	12	0.132	Same
	5	12	0.132	Same
	6	9	0.588	Same
	7	10	0.412	Same
	8	11	0.252	Same
	9	10	0.412	Same
	10	8	0.748	Same
	15	8	0.748	Same
	20	4	0.994	EV

Table A.89: Paired-t Comparison - EV vs. RM2 - Design 6, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-1.9365	0.8588	Same
	3	-2.6664	0.3311	Same
	4	-1.6459	1.5421	Same
	5	-2.8964	-0.0445	RM2
	6	-2.3559	0.6918	Same
	7	-2.1620	1.0666	Same
	8	-0.4005	2.3760	Same
	9	-0.2384	2.7178	Same
	10	-2.7006	0.5979	Same
	15	-1.7022	1.6818	Same
	20	-2.3368	0.9158	Same
SP	2	-1.8415	-0.7975	RM2
	3	-1.0651	-0.2740	RM2
	4	-0.6548	0.0738	Same
	5	-0.5265	0.1640	Same
	6	-0.3595	0.2866	Same
	7	-0.3121	0.2746	Same
	8	-0.3261	0.2222	Same
	9	-0.2392	0.3151	Same
	10	-0.1642	0.3062	Same
	15	0.0401	0.3650	EV
	20	0.1321	0.3970	EV

Table A.90: Sign Test Comparison - EV vs. RM2 - Design 6, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	11	0.252	Same
	3	13	0.058	RM2
	4	10	0.412	Same
	5	14	0.021	RM2
	6	11	0.252	Same
	7	11	0.252	Same
	8	7	0.868	Same
	9	6	0.942	EV
	10	13	0.058	RM2
	15	10	0.412	Same
	20	12	0.132	Same
SP	2	17	0.000	RM2
	3	16	0.001	RM2
	4	12	0.132	Same
	5	12	0.132	Same
	6	9	0.588	Same
	7	10	0.412	Same
	8	11	0.252	Same
	9	10	0.412	Same
	10	8	0.748	Same
	15	8	0.748	Same
	20	4	0.994	EV

Table A.91: Paired-t Comparison - EV vs. RM2 - Design 7, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-0.7353	0.8874	Same
	3	-0.6264	0.9886	Same
	4	-0.5683	0.9581	Same
	5	-0.1833	1.4900	Same
	6	0.4184	1.6724	EV
	7	-0.6824	1.0609	Same
	8	-0.0692	1.5305	Same
	9	0.1155	1.5205	EV
	10	-0.8252	0.8718	Same
	15	0.2590	1.6459	EV
	20	0.1585	1.6698	EV
SP	2	-1.2173	1.0422	Same
	3	-0.0350	1.6228	Same
	4	0.7992	1.8786	EV
	5	0.4902	1.6374	EV
	6	1.1301	1.8454	EV
	7	1.0325	1.7683	EV
	8	1.0656	1.6650	EV
	9	1.1106	1.6153	EV
	10	1.1316	1.6921	EV
	15	1.3557	1.7890	EV
	20	1.3239	1.7584	EV

Table A.92: Sign Test Comparison - EV vs. RM2 - Design 7, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	11	0.252	Same
	3	10	0.412	Same
	4	8	0.748	Same
	5	7	0.868	Same
	6	4	0.994	EV
	7	9	0.588	Same
	8	6	0.942	EV
	9	6	0.942	EV
	10	10	0.412	Same
	15	5	0.979	EV
	20	6	0.942	EV
SP	2	9	0.588	Same
	3	5	0.979	EV
	4	2	1.000	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	2	1.000	EV
	8	1	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.93: Paired-t Comparison - EV vs. RM2 - Design 7, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-5.5318	30.6017	Same
	3	-7.0840	27.0558	Same
	4	8.7173	36.2265	EV
	5	13.1518	43.3304	EV
	6	21.9651	40.9558	EV
	7	9.8734	33.4534	EV
	8	12.5963	39.7359	EV
	9	28.2522	46.8092	EV
	10	18.3721	43.0553	EV
	15	31.0077	45.4862	EV
	20	31.3738	48.5031	EV
SP	2	-30.4323	26.0557	Same
	3	-0.8758	40.5688	Same
	4	19.9788	46.9646	EV
	5	12.2540	40.9343	EV
	6	28.2514	46.1348	EV
	7	25.8134	44.2065	EV
	8	26.6408	41.6260	EV
	9	27.7653	40.3819	EV
	10	28.2897	42.3035	EV
	15	33.8924	44.7259	EV
	20	33.0973	43.9612	EV

Table A.94: Sign Test Comparison - EV vs. RM2 - Design 7, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	8	0.748	Same
	3	8	0.748	Same
	4	4	0.994	EV
	5	4	0.994	EV
	6	3	0.999	EV
	7	5	0.979	EV
	8	3	0.999	EV
	9	0	1.000	EV
	10	4	0.994	EV
	15	0	1.000	EV
	20	2	1.000	EV
SP	2	9	0.588	Same
	3	5	0.979	EV
	4	2	1.000	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	2	1.000	EV
	8	1	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.95: Paired-t Comparison - EV vs. RM2 - Design 7, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-1.0849	0.7233	Same
	3	-0.9016	0.9971	Same
	4	-0.8991	0.7328	Same
	5	-0.9181	1.0214	Same
	6	-0.9440	0.7616	Same
	7	-0.6203	1.2116	Same
	8	-0.3654	1.3028	Same
	9	-0.8520	1.0056	Same
	10	-1.6697	0.0675	Same
	15	-0.9131	0.9316	Same
	20	-0.6489	1.2813	Same
SP	2	-1.2261	1.0466	Same
	3	-0.0427	1.6270	Same
	4	0.8023	1.8776	EV
	5	0.4746	1.6409	EV
	6	1.1390	1.8531	EV
	7	1.0256	1.7680	EV
	8	1.0694	1.6599	EV
	9	1.1105	1.6161	EV
	10	1.1438	1.7029	EV
	15	1.3487	1.7854	EV
	20	1.3225	1.7618	EV

Table A.96: Sign Test Comparison - EV vs. RM2 - Design 7, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	12	0.132	Same
	3	11	0.252	Same
	4	9	0.588	Same
	5	9	0.588	Same
	6	10	0.412	Same
	7	9	0.588	Same
	8	7	0.868	Same
	9	10	0.412	Same
	10	14	0.021	RM2
	15	10	0.412	Same
	20	9	0.588	Same
SP	2	9	0.588	Same
	3	5	0.979	EV
	4	2	1.000	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	2	1.000	EV
	8	1	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.97: Paired-t Comparison - EV vs. RM2 - Design 8, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-1.0803	1.4406	Same
	3	-0.8098	1.6047	Same
	4	-1.1793	1.4754	Same
	5	-0.4996	2.0233	Same
	6	-1.2104	1.2079	Same
	7	0.4443	2.7243	EV
	8	0.1100	2.4413	EV
	9	-0.1996	2.2434	Same
	10	0.5869	2.7020	EV
	15	-0.8706	1.5435	Same
	20	0.1107	2.3516	EV
SP	2	-2.2222	-0.8934	RM2
	3	-1.3280	-0.0083	RM2
	4	-0.6540	0.3371	Same
	5	-0.2149	0.5738	Same
	6	-0.4345	0.4150	Same
	7	-0.0310	0.6501	Same
	8	0.0467	0.7032	EV
	9	0.1929	0.7252	EV
	10	0.2630	0.7159	EV
	15	0.4012	0.7657	EV
	20	0.4826	0.8797	EV

Table A.98: Sign Test Comparison - EV vs. RM2 - Design 8, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	9	0.588	Same
	3	9	0.588	Same
	4	9	0.588	Same
	5	8	0.748	Same
	6	10	0.412	Same
	7	4	0.994	EV
	8	6	0.942	EV
	9	7	0.868	Same
	10	5	0.979	EV
	15	9	0.588	Same
	20	6	0.942	EV
SP	2	17	0.000	RM2
	3	13	0.058	RM2
	4	11	0.252	Same
	5	7	0.868	Same
	6	10	0.412	Same
	7	6	0.942	EV
	8	7	0.868	Same
	9	5	0.979	EV
	10	4	0.994	EV
	15	2	1.000	EV
	20	1	1.000	EV

Table A.99: Paired-t Comparison - EV vs. RM2 - Design 8, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	2.6834	48.1846	EV
	3	15.5554	62.3135	EV
	4	35.5256	65.8021	EV
	5	38.2403	68.5441	EV
	6	41.7532	67.3401	EV
	7	50.6588	72.8480	EV
	8	53.8461	75.0428	EV
	9	44.2590	68.4258	EV
	10	53.5695	74.5905	EV
	15	49.4530	71.4416	EV
	20	55.1846	72.7830	EV
SP	2	-55.4672	-22.4214	RM2
	3	-33.4622	-0.4163	RM2
	4	-16.3722	8.4412	Same
	5	-5.4527	14.3544	Same
	6	-10.8336	10.3710	Same
	7	-0.7882	16.2052	Same
	8	1.1095	17.5657	EV
	9	4.8328	18.2988	EV
	10	6.5649	17.9138	EV
	15	10.0101	19.1493	EV
	20	11.9621	21.8995	EV

Table A.100: Sign Test Comparison - EV vs. RM2 - Design 8, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	5	0.979	EV
	3	5	0.979	EV
	4	2	1.000	EV
	5	1	1.000	EV
	6	0	1.000	EV
	7	1	1.000	EV
	8	1	1.000	EV
	9	0	1.000	EV
	10	1	1.000	EV
	15	1	1.000	EV
	20	0	1.000	EV
SP	2	17	0.000	RM2
	3	13	0.058	RM2
	4	11	0.252	Same
	5	7	0.868	Same
	6	10	0.412	Same
	7	6	0.942	EV
	8	7	0.868	Same
	9	5	0.979	EV
	10	4	0.994	EV
	15	2	1.000	EV
	20	2	1.000	EV

Table A.101: Paired-t Comparison - EV vs. RM2 - Design 8, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-1.1506	1.4068	Same
	3	-1.7278	0.9731	Same
	4	-1.1144	1.6113	Same
	5	-1.5223	1.2030	Same
	6	-2.0678	0.8129	Same
	7	-0.5835	2.1685	Same
	8	-1.1848	1.5744	Same
	9	-1.1775	1.7202	Same
	10	-1.1777	1.5377	Same
	15	-2.6864	-0.2029	RM2
	20	-0.9843	1.7362	Same
SP	2	-2.2387	-0.9130	RM2
	3	-1.3264	-0.0010	RM2
	4	-0.6560	0.3361	Same
	5	-0.2165	0.5711	Same
	6	-0.4265	0.4205	Same
	7	-0.0325	0.6448	Same
	8	0.0459	0.7043	EV
	9	0.1977	0.7345	EV
	10	0.2628	0.7145	EV
	15	0.4041	0.7656	EV
	20	0.4783	0.8785	EV

Table A.102: Sign Test Comparison - EV vs. RM2 - Design 8, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	9	0.588	Same
	3	11	0.252	Same
	4	9	0.588	Same
	5	11	0.252	Same
	6	12	0.132	Same
	7	7	0.868	Same
	8	9	0.588	Same
	9	9	0.588	Same
	10	10	0.412	Same
	15	15	0.006	RM2
	20	9	0.588	Same
SP	2	17	0.000	RM2
	3	13	0.058	RM2
	4	11	0.252	Same
	5	7	0.868	Same
	6	10	0.412	Same
	7	7	0.868	EV
	8	7	0.868	Same
	9	5	0.979	EV
	10	4	0.994	EV
	15	2	1.000	EV
	20	2	1.000	EV

Table A.103: Paired-t Comparison - EV vs. RM2 - Design 9, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	0.2449	3.0197	EV
	3	-1.0973	2.1124	Same
	4	0.0181	2.9251	EV
	5	-0.8223	2.1263	Same
	6	-0.4545	2.3602	Same
	7	0.4735	3.1315	EV
	8	0.1248	2.8280	EV
	9	-0.0840	2.7176	Same
	10	-0.8282	2.2183	Same
	15	-0.8031	2.3675	Same
	20	-0.1999	2.5945	Same
SP	2	-1.7952	-1.4377	RM2
	3	-1.1542	-0.8190	RM2
	4	-0.8534	-0.3344	RM2
	5	-0.7013	-0.2298	RM2
	6	-0.6061	-0.1403	RM2
	7	-0.3462	0.0957	Same
	8	-0.3198	0.1251	Same
	9	-0.1860	0.2143	Same
	10	-0.1860	0.2013	Same
	15	-0.0514	0.3024	Same
	20	0.0379	0.3093	EV

Table A.104: Sign Test Comparison - EV vs. RM2 - Design 9, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	5	0.979	EV
	3	9	0.588	Same
	4	6	0.942	EV
	5	10	0.412	Same
	6	8	0.748	Same
	7	5	0.979	EV
	8	5	0.979	EV
	9	6	0.942	EV
	10	8	0.748	Same
	15	8	0.748	RM2
	20	6	0.942	EV
SP	2	20	0.000	RM2
	3	20	0.000	RM2
	4	15	0.006	RM2
	5	16	0.001	RM2
	6	15	0.006	RM2
	7	12	0.132	Same
	8	10	0.412	Same
	9	10	0.412	Same
	10	11	0.252	Same
	15	9	0.588	Same
	20	5	0.979	EV

Table A.105: Paired-t Comparison - EV vs. RM2 - Design 9, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	45.1704	88.6245	EV
	3	43.4981	85.5999	EV
	4	70.0015	90.1412	EV
	5	58.8030	86.3427	EV
	6	62.6216	83.6273	EV
	7	71.9673	87.0381	EV
	8	48.5658	76.5549	EV
	9	57.7390	82.3789	EV
	10	60.9790	86.5027	EV
	15	55.7473	79.6389	EV
	20	67.0472	84.9386	EV
SP	2	-44.8204	-35.9821	RM2
	3	-28.8316	-20.5248	RM2
	4	-21.4543	-8.4116	RM2
	5	-17.4528	-5.6668	RM2
	6	-15.1783	-3.4019	RM2
	7	-8.6741	2.3808	Same
	8	-7.9238	3.1110	Same
	9	-4.5702	5.4055	Same
	10	-4.6998	4.9477	Same
	15	-1.2570	7.5890	Same
	20	0.9183	7.6325	EV

Table A.106: Sign Test Comparison - EV vs. RM2 - Design 9, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	3	0.999	EV
	4	1	1.000	EV
	5	1	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	1	1.000	EV
	9	1	1.000	EV
	10	1	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	20	0.000	RM2
	3	20	0.000	RM2
	4	15	0.006	RM2
	5	16	0.001	RM2
	6	15	0.006	RM2
	7	12	0.132	Same
	8	10	0.412	Same
	9	10	0.412	Same
	10	11	0.252	Same
	15	8	0.748	Same
	20	5	0.979	EV

Table A.107: Paired-t Comparison - EV vs. RM2 - Design 9, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-0.0292	3.2139	Same
	3	-1.0740	2.4414	Same
	4	-1.1236	2.4192	Same
	5	-0.7360	2.6381	Same
	6	-1.7474	1.4701	Same
	7	0.0963	2.8639	EV
	8	-0.0631	2.8150	Same
	9	-1.4091	1.9346	Same
	10	-2.3435	0.8179	Same
	15	-1.6780	1.6602	Same
	20	-1.6534	1.7686	Same
SP	2	-1.7853	-1.4344	RM2
	3	-1.1462	-0.8192	RM2
	4	-0.8550	-0.3299	RM2
	5	-0.7003	-0.2298	RM2
	6	-0.6087	-0.1379	RM2
	7	-0.3509	0.0945	Same
	8	-0.3198	0.1236	Same
	9	-0.1822	0.2196	Same
	10	-0.1885	0.1988	Same
	15	-0.0523	0.3026	Same
	20	0.0355	0.3062	EV

Table A.108: Sign Test Comparison - EV vs. RM2 - Design 9, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	7	0.868	Same
	3	8	0.748	Same
	4	8	0.748	Same
	5	7	0.868	Same
	6	11	0.252	Same
	7	6	0.942	EV
	8	6	0.942	EV
	9	10	0.412	Same
	10	12	0.132	Same
	15	10	0.412	Same
	20	10	0.412	Same
SP	2	20	0.000	RM2
	3	20	0.000	RM2
	4	15	0.006	RM2
	5	16	0.001	RM2
	6	15	0.006	RM2
	7	12	0.132	Same
	8	10	0.412	Same
	9	10	0.412	Same
	10	11	0.252	Same
	15	9	0.588	Same
	20	5	0.979	EV

Table A.109: Paired-t Comparison - EV vs. RM3 - Design 1, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-9.9072	0.3964	Same
	3	-6.6080	-1.2627	RM3
	4	-7.6236	-1.9270	RM3
	5	-1.7612	0.1644	Same
	6	-3.6657	-0.4789	RM3
	7	-3.3040	-0.7138	RM3
	8	-3.0568	-0.0819	RM3
	9	-3.6828	-0.4237	RM3
	10	-1.7624	0.1599	Same
	15	-2.0377	0.3038	Same
	20	-2.0176	-0.0752	RM3
SP	2	0.6310	1.6569	EV
	3	0.2400	1.7600	EV
	4	1.0775	1.9217	EV
	5	0.9044	1.8166	EV
	6	0.6794	1.6561	EV
	7	0.9093	1.8506	EV
	8	0.5461	1.5698	EV
	9	0.0828	1.7337	EV
	10	0.3876	1.6683	EV
	15	0.6380	1.7720	EV
	20	0.9957	1.7740	EV

Table A.110: Sign Test Comparison - EV vs. RM3 - Design 1, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	12	0.132	Same
	3	17	0.000	RM3
	4	16	0.001	RM3
	5	12	0.132	Same
	6	14	0.021	RM3
	7	15	0.006	RM3
	8	14	0.021	RM3
	9	14	0.021	RM3
	10	13	0.058	RM3
	15	13	0.058	RM3
	20	14	0.021	RM3
SP	2	1	1.000	EV
	3	3	0.999	EV
	4	2	1.000	EV
	5	3	0.999	EV
	6	1	1.000	EV
	7	1	1.000	EV
	8	2	1.000	EV
	9	3	0.999	EV
	10	2	1.000	EV
	15	3	0.999	EV
	20	2	1.000	EV

Table A.111: Paired-t Comparison - EV vs. RM3 - Design 1, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-100.1545	-11.7240	RM3
	3	-56.2738	-15.0710	RM3
	4	-69.7889	-17.9848	RM3
	5	-43.0398	-7.7881	RM3
	6	-44.2727	-9.8620	RM3
	7	-36.6524	-7.7766	RM3
	8	-25.6075	-8.4985	RM3
	9	-29.7029	9.4903	Same
	10	-26.3542	10.1345	Same
	15	-9.5566	21.7752	Same
	20	6.1495	35.1513	EV
SP	2	16.0053	41.3329	EV
	3	6.1304	43.9008	EV
	4	26.8484	47.9194	EV
	5	22.7088	45.5617	EV
	6	17.2967	41.3612	EV
	7	22.5337	46.0010	EV
	8	13.4548	39.3811	EV
	9	2.3321	43.3192	EV
	10	9.4072	41.5769	EV
	15	16.1020	44.3709	EV
	20	25.0930	44.4554	EV

Table A.112: Sign Test Comparison - EV vs. RM3 - Design 1, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	14	0.021	RM3
	3	16	0.001	RM3
	4	17	0.000	RM3
	5	18	0.000	RM3
	6	15	0.006	RM3
	7	17	0.000	RM3
	8	17	0.000	RM3
	9	13	0.058	RM3
	10	14	0.021	RM3
	15	8	0.748	RM3
	20	4	0.994	RM3
SP	2	1	1.000	EV
	3	3	0.999	EV
	4	2	1.000	EV
	5	2	1.000	EV
	6	1	1.000	EV
	7	1	1.000	EV
	8	2	1.000	EV
	9	3	0.999	EV
	10	2	1.000	EV
	15	3	0.999	EV
	20	1	1.000	EV

Table A.113: Paired-t Comparison - EV vs. RM3 - Design 1, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-195.7034	4.2841	Same
	3	-82.3572	0.5898	Same
	4	-111.1852	4.9439	Same
	5	-31.7445	28.4454	Same
	6	-44.3662	8.7341	Same
	7	-37.0995	16.9197	Same
	8	-47.3525	1.6874	Same
	9	-40.1603	11.4204	Same
	10	-17.0971	20.1917	Same
	15	-32.1620	6.3672	Same
	20	-35.1093	-1.4574	RM3
SP	2	0.6373	1.6531	EV
	3	0.2521	1.7712	EV
	4	1.0836	1.9229	EV
	5	0.9253	1.8228	EV
	6	0.6821	1.6591	EV
	7	0.9019	1.8424	EV
	8	0.5462	1.5746	EV
	9	0.0713	1.7253	EV
	10	0.3901	1.6587	EV
	15	0.6499	1.7682	EV
	20	1.0050	1.7794	EV

Table A.114: Sign Test Comparison - EV vs. RM3 - Design 1, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	15	0.006	RM3
	3	14	0.021	RM3
	4	11	0.252	Same
	5	10	0.412	Same
	6	12	0.132	Same
	7	10	0.412	Same
	8	11	0.252	Same
	9	11	0.252	Same
	10	7	0.868	Same
	15	12	0.132	Same
	20	16	0.001	RM3
SP	2	1	1.000	EV
	3	3	0.999	EV
	4	2	1.000	EV
	5	3	0.999	EV
	6	1	1.000	EV
	7	1	1.000	EV
	8	3	0.999	EV
	9	3	0.999	EV
	10	2	1.000	EV
	15	3	0.999	EV
	20	1	1.000	EV

Table A.115: Paired-t Comparison - EV vs. RM3 - Design 2, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-10.0659	-1.3582	RM3
	3	-5.7008	1.8407	Same
	4	-5.4341	0.5936	Same
	5	-8.7496	-0.9160	RM3
	6	-7.5624	0.6178	Same
	7	-7.0870	-0.9071	RM3
	8	-8.3751	-0.7519	RM3
	9	-5.1513	-0.4904	RM3
	10	-3.2024	1.7548	Same
	15	-4.6262	-0.2408	RM3
	20	-1.2603	1.8064	Same
SP	2	0.0387	1.6445	EV
	3	0.1368	1.3528	EV
	4	0.2895	1.2115	EV
	5	0.7272	1.2324	EV
	6	0.3927	1.1871	EV
	7	0.7214	1.2223	EV
	8	0.6900	1.1620	EV
	9	0.2819	0.9298	EV
	10	0.6091	1.1171	EV
	15	0.4958	0.9733	EV
	20	0.5175	0.9497	EV

Table A.116: Sign Test Comparison - EV vs. RM3 - Design 2, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	15	0.006	RM3
	3	10	0.412	RM3
	4	14	0.021	RM3
	5	15	0.006	RM3
	6	14	0.021	RM3
	7	18	0.000	RM3
	8	16	0.001	RM3
	9	15	0.006	RM3
	10	11	0.252	Same
	15	13	0.058	RM3
	20	11	0.252	Same
SP	2	4	0.994	EV
	3	3	0.999	EV
	4	3	0.999	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	2	1.000	EV
	9	3	0.999	EV
	10	2	1.000	EV
	15	2	1.000	EV
	20	2	1.000	EV

Table A.117: Paired-t Comparison - EV vs. RM3 - Design 2, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-84.3435	-18.1779	RM3
	3	-48.8118	-19.0530	RM3
	4	-43.1620	-8.5405	RM3
	5	-42.4273	-2.7633	RM3
	6	-26.5623	18.6310	Same
	7	-14.1787	33.8160	Same
	8	-25.5035	18.6339	Same
	9	-24.6080	34.1639	Same
	10	-8.2608	38.0422	Same
	15	4.9959	42.4593	EV
	20	30.3858	54.7191	EV
SP	2	1.3027	40.9840	EV
	3	3.3805	33.8171	EV
	4	6.9595	30.2607	EV
	5	18.3991	30.9641	EV
	6	9.8926	29.6614	EV
	7	18.2913	30.4998	EV
	8	17.3420	29.0587	EV
	9	7.0364	23.1781	EV
	10	15.3027	27.8610	EV
	15	12.2990	24.3107	EV
	20	12.8987	23.7931	EV

Table A.118: Sign Test Comparison - EV vs. RM3 - Design 2, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	14	0.021	RM3
	3	16	0.001	RM3
	4	15	0.006	RM3
	5	12	0.132	Same
	6	11	0.252	Same
	7	7	0.868	Same
	8	9	0.588	Same
	9	10	0.412	Same
	10	7	0.868	Same
	15	7	0.868	Same
	20	1	1.000	EV
SP	2	4	0.994	EV
	3	3	0.999	EV
	4	3	0.999	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	2	1.000	EV
	9	3	0.999	EV
	10	2	1.000	EV
	15	2	1.000	EV
	20	2	1.000	EV

Table A.119: Paired-t Comparison - EV vs. RM3 - Design 2, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-142.1848	31.8308	Same
	3	-93.9895	29.7491	Same
	4	-39.0919	85.1953	Same
	5	-105.3847	55.2587	Same
	6	-119.3100	37.6914	Same
	7	-151.7668	-32.3811	RM3
	8	-106.6851	35.6610	Same
	9	-66.2646	4.8585	Same
	10	-112.7719	9.7025	Same
	15	-97.8959	-4.9254	RM3
	20	-39.6728	22.7429	EV
SP	2	0.0342	1.6412	EV
	3	0.1448	1.3541	EV
	4	0.2820	1.2177	EV
	5	0.7315	1.2382	EV
	6	0.3904	1.1806	EV
	7	0.7317	1.2209	EV
	8	0.6969	1.1602	EV
	9	0.2867	0.9302	EV
	10	0.6084	1.1076	EV
	15	0.4993	0.9769	EV
	20	0.5213	0.9589	EV

Table A.120: Sign Test Comparison - EV vs. RM3 - Design 2, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	13	0.058	RM3
	3	10	0.412	Same
	4	6	0.942	EV
	5	10	0.412	Same
	6	13	0.058	RM3
	7	15	0.006	RM3
	8	11	0.252	Same
	9	14	0.021	RM3
	10	12	0.132	Same
	15	14	0.021	RM3
	20	9	0.588	Same
SP	2	4	0.994	EV
	3	3	0.999	EV
	4	3	0.999	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	2	1.000	EV
	9	3	0.999	EV
	10	2	1.000	EV
	15	2	1.000	EV
	20	2	1.000	EV

Table A.121: Paired-t Comparison - EV vs. RM3 - Design 3, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-17.7868	7.1878	Same
	3	-29.6546	-0.9426	RM3
	4	-14.9798	1.0688	Same
	5	-9.3412	3.1988	Same
	6	-8.4517	2.5173	Same
	7	-10.3503	-2.4828	RM3
	8	-18.1567	-3.1759	RM3
	9	-10.3841	-0.0451	RM3
	10	-9.4833	-0.8133	RM3
	15	-6.4613	-0.7752	RM3
	20	-5.0095	0.8361	Same
SP	2	-0.6711	0.7778	Same
	3	-0.1750	0.7147	Same
	4	-0.1883	0.6510	Same
	5	-0.2347	0.4707	Same
	6	-0.1463	0.5751	Same
	7	-0.0444	0.6165	Same
	8	-0.1607	0.4647	Same
	9	-0.1639	0.4833	Same
	10	-0.0506	0.4304	Same
	15	-0.0792	0.3771	Same
	20	0.0603	0.4065	EV

Table A.122: Sign Test Comparison - EV vs. RM3 - Design 3, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	11	0.252	Same
	3	14	0.021	RM3
	4	12	0.132	Same
	5	14	0.021	RM3
	6	13	0.058	RM3
	7	15	0.006	RM3
	8	17	0.000	RM3
	9	13	0.058	RM3
	10	13	0.058	RM3
	15	14	0.021	RM3
	20	11	0.252	Same
SP	2	8	0.748	Same
	3	8	0.748	Same
	4	6	0.942	EV
	5	9	0.588	Same
	6	7	0.868	Same
	7	6	0.942	EV
	8	8	0.748	Same
	9	8	0.748	Same
	10	7	0.868	Same
	15	7	0.868	Same
	20	6	0.942	EV

Table A.123: Paired-t Comparison - EV vs. RM3 - Design 3, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-61.9295	-18.3157	RM3
	3	-91.8303	20.9417	Same
	4	-96.0959	-22.7217	RM3
	5	-27.5078	45.4550	Same
	6	-16.9336	16.7426	Same
	7	-12.8592	42.3940	Same
	8	-16.1783	38.3576	Same
	9	5.0165	47.7076	EV
	10	0.1672	43.8132	EV
	15	25.5128	56.9690	EV
	20	36.4193	62.7863	EV
SP	2	-16.8795	19.6389	Same
	3	-4.4923	17.8663	Same
	4	-4.7018	16.1304	Same
	5	-6.0702	11.8018	Same
	6	-3.9685	14.0907	Same
	7	-1.1366	15.4289	Same
	8	-4.2771	11.3675	Same
	9	-3.9512	12.1085	Same
	10	-1.4785	10.6279	Same
	15	-1.9217	9.5091	Same
	20	1.4479	10.0819	EV

Table A.124: Sign Test Comparison - EV vs. RM3 - Design 3, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	16	0.001	RM3
	3	14	0.021	RM3
	4	15	0.006	RM3
	5	10	0.412	Same
	6	10	0.412	Same
	7	7	0.868	Same
	8	8	0.748	Same
	9	6	0.942	EV
	10	5	0.979	EV
	15	2	1.000	EV
	20	1	1.000	EV
SP	2	7	0.868	Same
	3	8	0.748	Same
	4	6	0.942	EV
	5	8	0.748	Same
	6	8	0.748	Same
	7	6	0.942	EV
	8	8	0.748	Same
	9	8	0.748	Same
	10	7	0.868	Same
	15	7	0.868	Same
	20	6	0.942	EV

Table A.125: Paired-t Comparison - EV vs. RM3 - Design 3, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-386.1131	156.4717	Same
	3	-549.7079	-27.6540	RM3
	4	-246.9631	62.4314	Same
	5	-158.7375	53.1557	Same
	6	-137.2379	115.2459	Same
	7	-199.1559	-35.7545	RM3
	8	-370.5395	-43.5758	RM3
	9	-158.8850	62.4621	Same
	10	-197.9268	-4.3205	RM3
	15	-89.3709	34.0017	Same
	20	-110.6957	18.8716	Same
SP	2	-0.6785	0.7743	Same
	3	-0.1850	0.7087	Same
	4	-0.1842	0.6467	Same
	5	-0.2386	0.4746	Same
	6	-0.1643	0.5610	Same
	7	-0.0460	0.6213	Same
	8	-0.1669	0.4587	Same
	9	-0.1585	0.4840	Same
	10	-0.0585	0.4272	Same
	15	-0.0830	0.3749	Same
	20	0.0610	0.4020	EV

Table A.126: Sign Test Comparison - EV vs. RM3 - Design 3, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	9	0.588	Same
	3	14	0.021	RM3
	4	11	0.252	Same
	5	9	0.588	Same
	6	7	0.868	Same
	7	15	0.006	RM3
	8	13	0.058	RM3
	9	11	0.252	Same
	10	13	0.058	RM3
	15	11	0.252	Same
	20	12	0.132	Same
SP	2	8	0.748	Same
	3	8	0.748	Same
	4	6	0.942	EV
	5	8	0.748	Same
	6	8	0.748	Same
	7	5	0.979	EV
	8	8	0.748	Same
	9	8	0.748	Same
	10	8	0.748	Same
	15	8	0.748	Same
	20	6	0.942	EV

Table A.127: Paired-t Comparison - EV vs. RM3 - Design 4, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-2.3640	3.4447	Same
	3	1.3791	3.9187	EV
	4	-0.0540	3.4031	Same
	5	1.7683	3.3660	EV
	6	-0.0081	2.9450	Same
	7	0.3510	3.1834	EV
	8	-0.5661	2.2734	Same
	9	0.8975	2.5359	EV
	10	-0.0006	2.3372	Same
	15	-0.1558	1.9202	Same
	20	1.2456	2.4756	EV
SP	2	-0.4731	1.1991	Same
	3	-0.3646	0.6937	Same
	4	-1.0305	0.2004	Same
	5	-1.0790	-0.2190	RM3
	6	-0.2326	0.5467	Same
	7	-0.3214	0.3289	Same
	8	-0.4451	0.5852	Same
	9	-0.5383	0.1848	Same
	10	-0.2758	0.5188	Same
	15	-0.1188	0.5546	Same
	20	-0.0080	0.4211	Same

Table A.128: Sign Test Comparison - EV vs. RM3 - Design 4, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	4	0.994	EV
	4	3	0.999	EV
	5	2	1.000	EV
	6	3	0.999	EV
	7	4	0.994	EV
	8	6	0.942	EV
	9	3	0.999	EV
	10	4	0.994	EV
	15	3	0.999	EV
	20	2	1.000	EV
SP	2	9	0.588	Same
	3	10	0.412	Same
	4	13	0.058	RM3
	5	16	0.001	RM3
	6	8	0.748	Same
	7	9	0.588	Same
	8	10	0.412	Same
	9	14	0.021	RM3
	10	10	0.412	Same
	15	7	0.868	Same
	20	6	0.942	EV

Table A.129: Paired-t Comparison - EV vs. RM3 - Design 4, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-53.9086	7.3679	Same
	3	-26.5751	10.5307	Same
	4	-24.3889	5.1029	Same
	5	-14.5308	10.4739	Same
	6	-20.2888	1.5992	Same
	7	-16.5493	3.1884	Same
	8	-16.1695	0.8178	Same
	9	-25.3100	-4.7144	RM3
	10	-18.8495	-0.2043	RM3
	15	-13.9012	2.1046	Same
	20	-13.6001	-0.5936	RM3
SP	2	-2.4207	29.5678	Same
	3	9.3712	31.1315	EV
	4	-12.8609	16.3875	Same
	5	-10.2835	22.7044	Same
	6	-2.0169	22.7031	Same
	7	-1.2946	24.1962	Same
	8	4.4036	25.5402	EV
	9	-3.8916	19.2571	Same
	10	-3.4971	20.9100	Same
	15	4.7772	21.9503	EV
	20	6.8504	20.6213	EV

Table A.130: Sign Test Comparison - EV vs. RM3 - Design 4, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	11	0.252	Same
	3	11	0.252	Same
	4	12	0.132	Same
	5	11	0.252	Same
	6	13	0.058	RM3
	7	12	0.132	Same
	8	13	0.058	RM3
	9	14	0.021	RM3
	10	14	0.021	RM3
	15	12	0.132	Same
	20	15	0.006	RM3
SP	2	7	0.868	Same
	3	4	0.994	EV
	4	7	0.868	Same
	5	6	0.942	EV
	6	5	0.979	EV
	7	5	0.979	EV
	8	3	0.999	EV
	9	4	0.994	EV
	10	6	0.942	EV
	15	3	0.999	EV
	20	4	0.994	EV

Table A.131: Paired-t Comparison - EV vs. RM3 - Design 4, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-9.9136	76.4415	Same
	3	52.3259	87.8136	EV
	4	33.0198	77.3071	EV
	5	64.4996	90.9193	EV
	6	31.7541	74.6089	EV
	7	30.1311	74.8742	EV
	8	13.2045	61.9139	EV
	9	52.1312	80.4700	EV
	10	41.5152	77.8647	EV
	15	37.3468	70.9876	EV
	20	60.8326	80.0973	EV
SP	2	15.8510	41.8262	EV
	3	22.2263	42.4788	EV
	4	20.3525	38.3821	EV
	5	34.0206	42.5373	EV
	6	27.6416	45.3790	EV
	7	32.7860	51.3427	EV
	8	33.3813	54.4717	EV
	9	32.1655	42.8664	EV
	10	33.3617	46.6061	EV
	15	38.5761	47.6584	EV
	20	40.6718	45.6782	EV

Table A.132: Sign Test Comparison - EV vs. RM3 - Design 4, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	5	0.979	EV
	3	0	1.000	EV
	4	3	0.999	EV
	5	1	1.000	EV
	6	2	1.000	EV
	7	2	1.000	EV
	8	2	1.000	EV
	9	1	1.000	EV
	10	1	1.000	EV
	15	1	1.000	EV
	20	0	1.000	EV
SP	2	1	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.133: Paired-t Comparison - EV vs. RM3 - Design 5, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	3.9740	9.3900	EV
	3	1.0680	7.2079	EV
	4	-0.7957	6.7826	Same
	5	0.1473	7.8964	EV
	6	0.9834	6.1437	EV
	7	3.2128	7.7860	EV
	8	3.3143	6.0392	EV
	9	-0.1438	5.9958	Same
	10	4.0454	7.2795	EV
	15	2.5943	7.3133	EV
	20	4.2619	6.4903	EV
SP	2	-1.2610	0.1473	Same
	3	-0.0429	0.9236	Same
	4	0.0814	1.3357	EV
	5	0.2013	1.0886	EV
	6	0.2447	0.9495	EV
	7	0.2107	1.0189	EV
	8	0.3681	0.9383	EV
	9	0.2801	1.0361	EV
	10	0.5509	1.0942	EV
	15	0.8124	1.5108	EV
	20	0.8215	1.3516	EV

Table A.134: Sign Test Comparison - EV vs. RM3 - Design 5, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	3	0.999	EV
	4	5	0.979	EV
	5	3	0.999	EV
	6	4	0.994	EV
	7	3	0.999	EV
	8	2	1.000	EV
	9	6	0.942	EV
	10	2	1.000	EV
	15	2	1.000	EV
	20	1	1.000	EV
SP	2	10	0.412	Same
	3	8	0.748	Same
	4	8	0.748	Same
	5	4	0.994	EV
	6	2	1.000	EV
	7	5	0.979	EV
	8	3	0.999	EV
	9	3	0.999	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.135: Paired-t Comparison - EV vs. RM3 - Design 5, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-30.4079	20.9032	Same
	3	-34.5815	-2.0745	RM3
	4	-42.7816	-7.4093	RM3
	5	-39.7490	-6.5302	RM3
	6	-21.4710	-1.6698	RM3
	7	-16.4107	-4.8608	RM3
	8	-31.6309	-10.7005	RM3
	9	-18.5680	-1.4155	RM3
	10	-18.9348	-5.1630	RM3
	15	-21.9759	-4.9047	RM3
	20	-10.3828	4.6128	Same
SP	2	-30.0125	12.0694	Same
	3	-10.8842	14.1055	Same
	4	0.4974	12.6350	EV
	5	0.8734	16.9795	EV
	6	-12.2183	7.1879	Same
	7	-5.0090	11.0120	Same
	8	-5.0842	8.3165	Same
	9	-0.9386	10.9421	Same
	10	-3.1720	9.5439	Same
	15	-3.1639	7.3347	Same
	20	-0.9885	8.1819	Same

Table A.136: Sign Test Comparison - EV vs. RM3 - Design 5, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	8	0.748	Same
	3	13	0.058	RM3
	4	16	0.001	RM3
	5	15	0.006	RM3
	6	13	0.058	RM3
	7	15	0.006	RM3
	8	16	0.001	RM3
	9	13	0.058	RM3
	10	15	0.006	RM3
	15	16	0.001	RM3
	20	8	0.748	Same
SP	2	10	0.412	Same
	3	6	0.942	EV
	4	6	0.942	EV
	5	6	0.942	EV
	6	10	0.412	Same
	7	7	0.868	Same
	8	8	0.748	Same
	9	6	0.942	EV
	10	7	0.868	Same
	15	9	0.588	Same
	20	7	0.868	Same

Table A.137: Paired-t Comparison - EV vs. RM3 - Design 5, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	142.4264	234.4599	EV
	3	116.5535	206.4314	EV
	4	40.2218	203.9649	EV
	5	89.5414	205.2140	EV
	6	117.3928	197.3548	EV
	7	122.2437	202.4102	EV
	8	151.9852	193.4475	EV
	9	52.5287	184.2053	EV
	10	118.5242	191.7015	EV
	15	111.9947	185.1332	EV
	20	144.9129	193.0472	EV
SP	2	18.6985	37.6865	EV
	3	29.9426	51.8270	EV
	4	32.3279	58.6559	EV
	5	37.6564	57.3510	EV
	6	37.0064	50.4916	EV
	7	34.6346	50.3475	EV
	8	35.8282	46.5306	EV
	9	35.5365	50.1966	EV
	10	38.3240	51.5276	EV
	15	43.7142	55.8013	EV
	20	43.9096	52.3322	EV

Table A.138: Sign Test Comparison - EV vs. RM3 - Design 5, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	1	1.000	EV
	3	1	1.000	EV
	4	2	1.000	EV
	5	1	1.000	EV
	6	0	1.000	EV
	7	1	1.000	EV
	8	0	1.000	EV
	9	3	0.999	EV
	10	0	1.000	EV
	15	2	1.000	EV
	20	0	1.000	EV
SP	2	2	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.139: Paired-t Comparison - EV vs. RM3 - Design 6, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-3.6240	15.2672	Same
	3	5.6348	14.0677	EV
	4	2.6991	17.7173	EV
	5	8.6055	16.4795	EV
	6	5.7401	14.7220	EV
	7	5.8731	13.7986	EV
	8	4.5560	13.8063	EV
	9	8.2325	15.9564	EV
	10	9.9986	15.8048	EV
	15	10.4666	16.1276	EV
	20	12.1468	15.3757	EV
SP	2	-0.4956	1.3801	Same
	3	0.1619	1.1470	EV
	4	0.6407	2.1332	EV
	5	0.6153	1.3212	EV
	6	0.7379	1.6470	EV
	7	0.8286	1.6379	EV
	8	1.2298	2.0450	EV
	9	1.0190	1.6750	EV
	10	0.8920	1.5097	EV
	15	1.2149	1.7284	EV
	20	1.2145	1.6897	EV

Table A.140: Sign Test Comparison - EV vs. RM3 - Design 6, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	6	0.942	EV
	3	3	0.999	EV
	4	4	0.994	EV
	5	3	0.999	EV
	6	3	0.999	EV
	7	2	1.000	EV
	8	4	0.994	EV
	9	3	0.999	EV
	10	1	1.000	EV
	15	1	1.000	EV
	20	0	1.000	EV
SP	2	9	0.588	Same
	3	6	0.942	EV
	4	3	0.999	EV
	5	3	0.999	EV
	6	0	1.000	EV
	7	2	1.000	EV
	8	0	1.000	EV
	9	1	1.000	EV
	10	1	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.141: Paired-t Comparison - EV vs. RM3 - Design 6, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-88.8672	-1.0907	RM3
	3	-47.5464	-6.4914	RM3
	4	-67.3960	1.4201	Same
	5	-42.3117	-15.4863	RM3
	6	-40.7173	-19.6628	RM3
	7	-23.5521	0.8894	Same
	8	-40.3991	-11.9322	RM3
	9	-23.0045	-2.3129	RM3
	10	-25.1658	-8.5505	RM3
	15	-24.4841	-8.6987	RM3
	20	-15.2144	-3.3153	RM3
SP	2	-42.0241	-20.4878	RM3
	3	-22.5266	-8.1908	RM3
	4	-13.0411	-0.8555	RM3
	5	-10.9640	1.7315	Same
	6	-8.4968	1.8708	Same
	7	-8.1610	1.7071	Same
	8	-8.0269	-0.2702	RM3
	9	-6.2965	2.3402	Same
	10	-6.1664	1.4214	Same
	15	-2.8931	2.8808	Same
	20	-3.0219	2.1178	Same

Table A.142: Sign Test Comparison - EV vs. RM3 - Design 6, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	16	0.001	RM3
	3	10	0.412	Same
	4	16	0.001	RM3
	5	17	0.000	RM3
	6	17	0.000	RM3
	7	14	0.021	RM3
	8	15	0.006	RM3
	9	12	0.132	Same
	10	17	0.000	RM3
	15	18	0.000	RM3
	20	16	0.001	RM3
SP	2	17	0.000	RM3
	3	18	0.000	RM3
	4	14	0.021	RM3
	5	12	0.132	Same
	6	12	0.132	Same
	7	13	0.058	RM3
	8	14	0.021	RM3
	9	11	0.252	Same
	10	12	0.132	Same
	15	9	0.588	Same
	20	10	0.412	Same

Table A.143: Paired-t Comparison - EV vs. RM3 - Design 6, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-29.1338	344.0175	Same
	3	209.7298	409.4726	EV
	4	196.3094	466.5666	EV
	5	215.9664	429.7710	EV
	6	200.6343	410.0211	EV
	7	218.7235	387.2708	EV
	8	218.4136	387.2063	EV
	9	260.4510	436.7253	EV
	10	268.1545	414.1030	EV
	15	332.5130	437.9975	EV
	20	322.9677	412.9909	EV
SP	2	27.3470	60.0028	EV
	3	27.0710	47.4546	EV
	4	39.8292	66.8576	EV
	5	34.8432	50.6051	EV
	6	32.9600	52.8312	EV
	7	35.8829	53.6203	EV
	8	43.9760	60.9366	EV
	9	40.4636	56.0310	EV
	10	35.9122	48.6289	EV
	15	42.0084	53.2370	EV
	20	41.2632	51.3401	EV

Table A.144: Sign Test Comparison - EV vs. RM3 - Design 6, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	4	0.994	EV
	3	1	1.000	EV
	4	4	0.994	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	1	1.000	EV
	9	1	1.000	EV
	10	1	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.145: Paired-t Comparison - EV vs. RM3 - Design 7, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-2.1581	1.6617	Same
	3	-1.5975	0.4134	Same
	4	-1.6329	0.0626	Same
	5	-1.3443	-0.0102	RM3
	6	-0.6856	0.8040	Same
	7	-0.6181	0.4385	Same
	8	-0.8527	0.1186	Same
	9	-0.6343	0.3302	Same
	10	-0.9082	0.1984	Same
	15	-0.2560	0.9885	Same
	20	-0.1143	1.1473	Same
SP	2	-1.2184	1.0454	Same
	3	-0.0363	1.6311	Same
	4	0.7996	1.8812	EV
	5	0.4826	1.6401	EV
	6	1.1413	1.8504	EV
	7	1.0282	1.7665	EV
	8	1.0786	1.6662	EV
	9	1.1093	1.6162	EV
	10	1.1434	1.7028	EV
	15	1.3526	1.7903	EV
	20	1.3250	1.7606	EV

Table A.146: Sign Test Comparison - EV vs. RM3 - Design 7, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	8	0.748	Same
	3	9	0.588	Same
	4	12	0.132	Same
	5	14	0.021	RM3
	6	9	0.588	Same
	7	11	0.252	Same
	8	12	0.132	Same
	9	11	0.252	Same
	10	11	0.252	Same
	15	7	0.868	Same
	20	6	0.942	EV
SP	2	9	0.588	Same
	3	5	0.979	EV
	4	2	1.000	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	2	1.000	EV
	8	1	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.147: Paired-t Comparison - EV vs. RM3 - Design 7, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-22.6028	0.1091	Same
	3	-23.2581	-3.9275	RM3
	4	-6.3640	14.9309	Same
	5	0.0509	19.9747	EV
	6	3.9091	21.8313	EV
	7	0.8429	20.0609	EV
	8	2.7768	25.4966	EV
	9	17.8984	34.5096	EV
	10	9.5563	31.5892	EV
	15	25.2567	37.7035	EV
	20	26.6651	43.1499	EV
SP	2	-30.5175	25.9895	Same
	3	-0.5939	40.7037	Same
	4	20.1504	47.0337	EV
	5	12.2624	41.2380	EV
	6	28.4208	46.3515	EV
	7	25.7020	44.1594	EV
	8	26.7851	41.5929	EV
	9	27.7828	40.4801	EV
	10	28.4935	42.4613	EV
	15	33.8910	44.6937	EV
	20	33.2127	44.0695	EV

Table A.148: Sign Test Comparison - EV vs. RM3 - Design 7, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	13	0.058	RM3
	3	16	0.001	RM3
	4	6	0.942	EV
	5	4	0.994	EV
	6	4	0.994	EV
	7	7	0.868	Same
	8	5	0.979	EV
	9	2	1.000	EV
	10	4	0.994	EV
	15	0	1.000	EV
	20	2	1.000	EV
SP	2	9	0.588	Same
	3	5	0.979	EV
	4	2	1.000	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	2	1.000	EV
	8	1	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.149: Paired-t Comparison - EV vs. RM3 - Design 7, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-17.3688	50.6295	Same
	3	-19.0806	35.4381	Same
	4	-31.8826	13.4559	Same
	5	-11.0873	21.8927	Same
	6	-13.4160	17.8075	Same
	7	-5.7064	18.0134	Same
	8	-17.8828	2.7730	Same
	9	-4.0358	14.0619	Same
	10	-3.7468	12.9148	Same
	15	-7.4808	3.3428	Same
	20	-6.2973	2.0156	Same
SP	2	-1.2163	1.0480	Same
	3	-0.0209	1.6320	Same
	4	0.7854	1.8771	EV
	5	0.4781	1.6358	EV
	6	1.1410	1.8504	EV
	7	1.0405	1.7713	EV
	8	1.0782	1.6693	EV
	9	1.1141	1.6187	EV
	10	1.1391	1.7005	EV
	15	1.3501	1.7863	EV
	20	1.3241	1.7627	EV

Table A.150: Sign Test Comparison - EV vs. RM3 - Design 7, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	5	0.979	EV
	3	6	0.942	EV
	4	10	0.412	Same
	5	6	0.942	EV
	6	8	0.748	Same
	7	6	0.942	EV
	8	13	0.058	RM3
	9	5	0.979	EV
	10	9	0.588	Same
	15	10	0.412	Same
	20	11	0.252	Same
SP	2	9	0.588	Same
	3	5	0.979	EV
	4	2	1.000	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	2	1.000	EV
	8	1	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.151: Paired-t Comparison - EV vs. RM3 - Design 8, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-8.2043	1.3995	Same
	3	-2.6993	1.1810	Same
	4	-2.4161	-0.6534	RM3
	5	-1.4540	0.1466	Same
	6	-1.8332	-0.1940	RM3
	7	-1.2261	0.3971	Same
	8	-0.9494	0.1775	Same
	9	-0.7623	0.7353	Same
	10	-0.4300	0.7701	Same
	15	-1.1363	0.6396	Same
	20	-0.2859	1.4898	Same
SP	2	-2.2303	-0.9007	RM3
	3	-1.3267	-0.0049	RM3
	4	-0.6531	0.3403	Same
	5	-0.2112	0.5741	Same
	6	-0.4289	0.4156	Same
	7	-0.0336	0.6496	Same
	8	0.0489	0.7023	EV
	9	0.1953	0.7336	EV
	10	0.2670	0.7191	EV
	15	0.3995	0.7648	EV
	20	0.4819	0.8790	EV

Table A.152: Sign Test Comparison - EV vs. RM3 - Design 8, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	9	0.588	Same
	3	10	0.412	Same
	4	17	0.000	RM3
	5	11	0.252	Same
	6	12	0.132	Same
	7	12	0.132	Same
	8	12	0.132	RM3
	9	9	0.588	Same
	10	11	0.252	Same
	15	11	0.252	Same
	20	6	0.942	EV
SP	2	17	0.000	RM3
	3	13	0.058	RM3
	4	11	0.252	Same
	5	7	0.868	Same
	6	10	0.412	Same
	7	6	0.942	EV
	8	7	0.868	Same
	9	5	0.979	EV
	10	4	0.994	EV
	15	2	1.000	EV
	20	1	1.000	EV

Table A.153: Paired-t Comparison - EV vs. RM3 - Design 8, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-45.1756	-13.8359	RM3
	3	-2.8320	31.1893	Same
	4	12.9874	33.1328	EV
	5	18.5089	43.7313	EV
	6	24.5748	46.9945	EV
	7	36.9742	57.3573	EV
	8	40.8679	56.8204	EV
	9	36.5371	58.1738	EV
	10	44.5827	65.2072	EV
	15	43.2772	62.4146	EV
	20	49.8078	66.3924	EV
SP	2	-55.6199	-22.5168	RM3
	3	-33.3364	-0.1710	RM3
	4	-16.4133	8.4921	Same
	5	-5.3962	14.3687	Same
	6	-10.6404	10.5370	Same
	7	-0.7783	16.2970	Same
	8	1.2687	17.5856	EV
	9	4.8191	18.2575	EV
	10	6.4748	17.9450	EV
	15	10.0362	19.1218	EV
	20	12.0196	22.0119	EV

Table A.154: Sign Test Comparison - EV vs. RM3 - Design 8, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	17	0.000	RM3
	3	6	0.942	EV
	4	2	1.000	EV
	5	4	0.994	EV
	6	3	0.999	EV
	7	1	1.000	EV
	8	0	1.000	EV
	9	1	1.000	EV
	10	1	1.000	EV
	15	1	1.000	EV
	20	0	1.000	EV
SP	2	17	0.000	RM3
	3	13	0.058	RM3
	4	11	0.252	Same
	5	7	0.868	Same
	6	10	0.412	Same
	7	6	0.942	EV
	8	7	0.868	Same
	9	5	0.979	EV
	10	4	0.994	EV
	15	2	1.000	EV
	20	1	1.000	EV

Table A.155: Paired-t Comparison - EV vs. RM3 - Design 8, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-136.1800	80.4395	Same
	3	-48.3562	65.7046	Same
	4	-4.3234	41.3574	Same
	5	-10.8084	45.8800	Same
	6	-3.3789	37.8401	Same
	7	-38.1358	5.6065	Same
	8	-7.0881	35.7590	Same
	9	-31.8824	4.1287	Same
	10	-23.2752	14.6372	Same
	15	-3.9300	15.6559	Same
	20	-14.1316	6.9532	Same
SP	2	-2.2366	-0.9062	RM3
	3	-1.3391	-0.0141	RM3
	4	-0.6557	0.3373	Same
	5	-0.2117	0.5795	Same
	6	-0.4264	0.4197	Same
	7	-0.0320	0.6534	Same
	8	0.0473	0.7023	EV
	9	0.1957	0.7288	EV
	10	0.2601	0.7157	EV
	15	0.3998	0.7658	EV
	20	0.4790	0.8782	EV

Table A.156: Sign Test Comparison - EV vs. RM3 - Design 8, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	8	0.748	Same
	3	6	0.942	EV
	4	8	0.748	Same
	5	5	0.979	EV
	6	6	0.942	EV
	7	14	0.021	RM3
	8	5	0.979	EV
	9	14	0.021	RM3
	10	10	0.412	Same
	15	6	0.942	EV
	20	11	0.252	Same
SP	2	17	0.000	RM3
	3	13	0.058	RM3
	4	11	0.252	Same
	5	7	0.868	Same
	6	10	0.412	Same
	7	7	0.868	Same
	8	7	0.868	Same
	9	5	0.979	EV
	10	4	0.994	EV
	15	2	1.000	EV
	20	1	1.000	EV

Table A.157: Paired-t Comparison - EV vs. RM3 - Design 9, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-13.9855	3.3336	Same
	3	-6.3217	2.3635	Same
	4	-8.4842	-1.5380	RM3
	5	-3.9722	0.2178	Same
	6	-3.0024	0.7579	Same
	7	-3.1530	-0.0202	RM3
	8	-0.8665	0.9157	Same
	9	-0.9626	0.4456	Same
	10	-1.1261	1.0118	Same
	15	-1.1012	0.8638	Same
	20	-0.6657	1.3233	Same
SP	2	-1.8025	-1.4393	RM3
	3	-1.1531	-0.8240	RM3
	4	-0.8536	-0.3286	RM3
	5	-0.6941	-0.2283	RM3
	6	-0.6068	-0.1406	RM3
	7	-0.3490	0.0962	Same
	8	-0.3192	0.1225	Same
	9	-0.1845	0.2150	Same
	10	-0.1918	0.1950	Same
	15	-0.0507	0.3033	Same
	20	0.0375	0.3082	EV

Table A.158: Sign Test Comparison - EV vs. RM3 - Design 9, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	11	0.252	Same
	3	9	0.588	EV
	4	13	0.058	RM3
	5	12	0.132	Same
	6	11	0.252	Same
	7	14	0.021	RM3
	8	12	0.132	Same
	9	11	0.252	Same
	10	11	0.252	Same
	15	10	0.412	Same
	20	8	0.748	Same
SP	2	20	0.000	RM3
	3	20	0.000	RM3
	4	15	0.006	RM3
	5	16	0.001	RM3
	6	15	0.006	RM3
	7	12	0.132	Same
	8	10	0.412	Same
	9	10	0.412	Same
	10	11	0.252	Same
	15	9	0.588	Same
	20	5	0.979	EV

Table A.159: Paired-t Comparison - EV vs. RM3 - Design 9, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-40.1041	4.4087	Same
	3	1.4658	34.5461	EV
	4	27.0508	49.7310	EV
	5	32.3127	56.1462	EV
	6	41.3646	60.8986	EV
	7	53.1077	67.6541	EV
	8	40.0975	66.0324	EV
	9	42.5706	67.2365	EV
	10	50.8648	74.2823	EV
	15	50.0378	71.9580	EV
	20	60.7133	76.6575	EV
SP	2	-44.3174	-35.6691	RM3
	3	-28.8391	-20.4961	RM3
	4	-21.2558	-8.2195	RM3
	5	-17.5072	-5.7222	RM3
	6	-15.1979	-3.4693	RM3
	7	-8.6006	2.4273	Same
	8	-7.9929	3.0247	Same
	9	-4.6668	5.3361	Same
	10	-4.7247	4.9080	Same
	15	-1.4022	7.5354	Same
	20	0.8770	7.6783	EV

Table A.160: Sign Test Comparison - EV vs. RM3 - Design 9, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	11	0.252	Same
	3	5	0.979	EV
	4	2	1.000	RM3
	5	1	1.000	Same
	6	0	1.000	Same
	7	0	1.000	RM3
	8	1	1.000	Same
	9	1	1.000	Same
	10	1	1.000	Same
	15	0	1.000	Same
	20	0	1.000	Same
SP	2	20	0.000	RM3
	3	20	0.000	RM3
	4	15	0.006	RM3
	5	16	0.001	RM3
	6	15	0.006	RM3
	7	12	0.132	Same
	8	10	0.412	Same
	9	10	0.412	Same
	10	11	0.252	Same
	15	8	0.748	Same
	20	5	0.979	EV

Table A.161: Paired-t Comparison - EV vs. RM3 - Design 9, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-244.4457	160.5642	Same
	3	-87.8952	122.0230	Same
	4	-171.9301	31.5472	Same
	5	-63.1875	55.8486	Same
	6	-16.0122	93.0399	Same
	7	-64.1738	38.5775	Same
	8	-32.6204	47.4015	Same
	9	-39.0492	38.8967	Same
	10	-42.2239	31.8921	Same
	15	-38.0200	14.5053	Same
	20	-9.0345	22.8852	Same
SP	2	-1.7943	-1.4349	RM3
	3	-1.1600	-0.8205	RM3
	4	-0.8549	-0.3318	RM3
	5	-0.7020	-0.2330	RM3
	6	-0.6073	-0.1361	RM3
	7	-0.3489	0.0921	Same
	8	-0.3183	0.1243	Same
	9	-0.1851	0.2157	Same
	10	-0.1899	0.1968	Same
	15	-0.0510	0.3042	Same
	20	0.0341	0.3063	EV

Table A.162: Sign Test Comparison - EV vs. RM3 - Design 9, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	7	0.868	Same
	3	5	0.979	EV
	4	12	0.132	Same
	5	9	0.588	Same
	6	6	0.942	Same
	7	10	0.412	Same
	8	8	0.748	Same
	9	8	0.748	Same
	10	10	0.412	Same
	15	11	0.252	Same
	20	8	0.748	Same
SP	2	20	0.000	RM3
	3	20	0.000	RM3
	4	15	0.006	RM3
	5	16	0.001	RM3
	6	15	0.006	RM3
	7	12	0.132	Same
	8	10	0.412	Same
	9	10	0.412	Same
	10	11	0.252	Same
	15	9	0.588	Same
	20	5	0.979	EV

Table A.163: Paired-t Comparison - EV vs. OM1 - Design 1, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-8.8258	2.0035	Same
	3	-4.4970	0.1969	Same
	4	-7.5889	0.9018	Same
	5	1.5492	3.4783	EV
	6	-0.4649	2.7857	Same
	7	-0.1511	2.8570	Same
	8	-0.3415	3.8375	Same
	9	-1.0436	2.5958	Same
	10	1.6611	3.6055	EV
	15	-0.7633	1.9539	Same
	20	-0.0679	2.9589	Same
SP	2	-0.9437	0.9449	Same
	3	-0.5108	1.6456	Same
	4	-0.2359	2.1236	Same
	5	-0.4492	1.0154	Same
	6	0.1576	1.6075	EV
	7	0.3343	1.7980	EV
	8	0.0814	1.8498	EV
	9	0.2639	1.9848	EV
	10	0.9090	2.1406	EV
	15	1.5737	3.4770	EV
	20	2.2351	3.4663	EV

Table A.164: Sign Test Comparison - EV vs. OM1 - Design 1, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	11	0.252	Same
	3	12	0.132	Same
	4	12	0.132	Same
	5	4	0.994	EV
	6	8	0.748	Same
	7	5	0.979	EV
	8	5	0.979	EV
	9	7	0.868	Same
	10	1	1.000	EV
	15	9	0.588	Same
	20	2	1.000	EV
SP	2	12	0.132	Same
	3	8	0.748	Same
	4	9	0.588	Same
	5	9	0.588	Same
	6	4	0.994	EV
	7	5	0.979	EV
	8	6	0.942	EV
	9	4	0.994	EV
	10	3	0.999	EV
	15	1	1.000	EV
	20	0	1.000	EV

Table A.165: Paired-t Comparison - EV vs. OM1 - Design 1, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-78.3095	5.4457	Same
	3	-55.2780	-3.5908	OM1
	4	-77.2300	6.6976	Same
	5	-45.9474	9.1081	Same
	6	-44.8769	3.9660	Same
	7	-22.0283	0.7576	Same
	8	-37.9253	-3.0784	OM1
	9	-34.1964	2.7412	Same
	10	-17.6705	8.0715	Same
	15	-23.9196	-3.8394	OM1
	20	-19.4634	-0.0364	OM1
SP	2	-10.1916	20.7636	Same
	3	-23.6077	12.2845	Same
	4	-3.9360	18.5193	Same
	5	-12.9728	12.5833	Same
	6	-19.6300	6.6849	Same
	7	-14.5473	7.6541	Same
	8	-23.3889	-0.4453	OM1
	9	-23.1538	2.9401	Same
	10	-19.5536	1.0286	Same
	15	-13.7882	1.2133	Same
	20	-7.3188	2.9303	Same

Table A.166: Sign Test Comparison - EV vs. OM1 - Design 1, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	12	0.132	Same
	3	14	0.021	OM1
	4	13	0.058	OM1
	5	9	0.588	Same
	6	12	0.132	Same
	7	11	0.252	Same
	8	14	0.021	OM1
	9	15	0.006	OM1
	10	8	0.748	Same
	15	14	0.021	OM1
	20	14	0.021	OM1
SP	2	13	0.058	OM1
	3	9	0.588	Same
	4	5	0.979	EV
	5	8	0.748	Same
	6	11	0.252	Same
	7	10	0.412	Same
	8	12	0.132	Same
	9	14	0.021	OM1
	10	12	0.132	Same
	15	13	0.058	OM1
	20	14	0.021	OM1

Table A.167: Paired-t Comparison - EV vs. OM1 - Design 1, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-163.0255	35.3560	Same
	3	-27.1374	47.2144	Same
	4	-70.7601	70.3111	Same
	5	28.8120	80.7113	EV
	6	26.6461	76.7835	EV
	7	26.0939	77.0860	EV
	8	25.9717	74.9956	EV
	9	28.0785	82.0808	EV
	10	58.2952	92.1462	EV
	15	32.2960	72.6446	EV
	20	37.5715	80.6007	EV
SP	2	22.6885	59.2627	EV
	3	39.1652	74.8114	EV
	4	58.0534	93.6321	EV
	5	50.4986	80.0794	EV
	6	62.9770	85.1848	EV
	7	61.8018	88.6929	EV
	8	64.8262	80.7840	EV
	9	63.0511	84.3088	EV
	10	76.9026	95.0991	EV
	15	75.9909	100.2541	EV
	20	81.9130	98.6266	EV

Table A.168: Sign Test Comparison - EV vs. OM1 - Design 1, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	11	0.252	Same
	3	5	0.979	EV
	4	7	0.868	Same
	5	3	0.999	EV
	6	5	0.979	EV
	7	3	0.999	EV
	8	2	1.000	EV
	9	1	1.000	EV
	10	1	1.000	EV
	15	1	1.000	EV
	20	1	1.000	EV
SP	2	1	1.000	EV
	3	1	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.169: Paired-t Comparison - EV vs. OM1 - Design 2, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-3.5444	4.0910	Same
	3	0.4039	7.3564	EV
	4	2.0723	7.7043	EV
	5	-1.9742	8.5462	Same
	6	-1.1812	7.8201	Same
	7	-5.4935	7.5133	Same
	8	-2.6134	8.2588	Same
	9	2.9777	7.8434	EV
	10	3.7196	9.2965	EV
	15	2.2718	8.5785	EV
	20	7.9624	10.4586	EV
SP	2	0.0862	2.0275	EV
	3	-0.0914	1.5726	Same
	4	-0.3520	1.1079	Same
	5	0.5448	2.2039	EV
	6	1.3937	3.3752	EV
	7	2.1218	4.6822	EV
	8	1.4158	3.2236	EV
	9	1.5649	3.1913	EV
	10	2.0356	3.8780	EV
	15	2.5660	3.9659	EV
	20	2.6122	3.6446	EV

Table A.170: Sign Test Comparison - EV vs. OM1 - Design 2, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	8	0.748	Same
	3	6	0.942	EV
	4	4	0.994	EV
	5	2	1.000	EV
	6	5	0.979	EV
	7	5	0.979	EV
	8	4	0.994	EV
	9	2	1.000	EV
	10	1	1.000	EV
	15	3	0.999	EV
	20	0	1.000	EV
SP	2	4	0.994	EV
	3	7	0.868	Same
	4	10	0.412	Same
	5	4	0.994	EV
	6	1	1.000	EV
	7	0	1.000	EV
	8	2	1.000	EV
	9	2	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.171: Paired-t Comparison - EV vs. OM1 - Design 2, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-54.3409	12.7188	Same
	3	-25.3665	5.6338	Same
	4	-33.3065	10.0086	Same
	5	-41.1849	6.0418	Same
	6	-43.7910	1.9295	Same
	7	-51.4110	7.4205	Same
	8	-50.9596	8.6397	Same
	9	-56.4635	-4.0295	OM1
	10	-38.0162	1.9777	EV
	15	-45.3891	-10.5209	OM1
	20	-21.2610	0.2970	EV
SP	2	-20.7178	15.8970	EV
	3	-9.4361	16.7404	Same
	4	-10.5880	8.7872	Same
	5	-2.3350	9.5555	EV
	6	-3.3375	6.7752	EV
	7	-4.3052	8.1924	EV
	8	-2.1049	7.7178	EV
	9	-5.1303	5.9070	EV
	10	-0.2623	7.7445	EV
	15	-6.6636	0.4909	EV
	20	-3.9484	2.4629	EV

Table A.172: Sign Test Comparison - EV vs. OM1 - Design 2, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	13	0.058	OM1
	3	12	0.132	Same
	4	13	0.058	OM1
	5	11	0.252	Same
	6	14	0.021	OM1
	7	9	0.588	Same
	8	12	0.132	Same
	9	15	0.006	OM1
	10	11	0.252	Same
	15	17	0.000	OM1
	20	14	0.021	OM1
SP	2	8	0.748	Same
	3	6	0.942	EV
	4	9	0.588	Same
	5	8	0.748	Same
	6	7	0.868	Same
	7	8	0.748	Same
	8	8	0.748	Same
	9	10	0.412	Same
	10	7	0.868	Same
	15	14	0.021	OM1
	20	10	0.412	Same

Table A.173: Paired-t Comparison - EV vs. OM1 - Design 2, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	28.9318	161.6761	EV
	3	38.6292	154.3778	EV
	4	130.3000	239.4215	EV
	5	66.6932	253.1714	EV
	6	92.6982	231.2785	EV
	7	-37.1287	184.3108	Same
	8	50.3712	250.0767	EV
	9	149.2779	228.6555	EV
	10	87.0565	225.4615	EV
	15	67.2897	234.8020	EV
	20	194.1491	259.3294	EV
SP	2	39.6553	77.0524	EV
	3	39.2059	83.5707	EV
	4	46.6823	68.6118	EV
	5	55.3386	86.1522	EV
	6	74.1421	104.4089	EV
	7	76.9356	120.6860	EV
	8	66.6747	102.4986	EV
	9	70.0930	93.2435	EV
	10	82.8581	109.6575	EV
	15	82.9829	114.1737	EV
	20	83.2527	101.4160	EV

Table A.174: Sign Test Comparison - EV vs. OM1 - Design 2, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	5	0.979	EV
	3	4	0.994	EV
	4	1	1.000	EV
	5	1	1.000	EV
	6	2	1.000	EV
	7	3	0.999	EV
	8	2	1.000	EV
	9	0	1.000	EV
	10	3	0.999	EV
	15	2	1.000	EV
	20	1	1.000	EV
SP	2	1	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.175: Paired-t Comparison - EV vs. OM1 - Design 3, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-5.1646	16.3151	Same
	3	-20.3493	14.4030	Same
	4	-0.1193	17.5752	Same
	5	9.0917	18.0090	EV
	6	7.8172	20.2111	EV
	7	9.8945	16.9443	Same
	8	0.4788	20.3533	EV
	9	11.3795	21.2475	EV
	10	13.0436	21.8250	EV
	15	18.2946	24.2037	EV
	20	18.4923	24.2822	EV
SP	2	-0.1037	1.3584	Same
	3	1.3128	4.0088	EV
	4	1.1473	3.1923	EV
	5	1.6906	3.8467	EV
	6	1.7107	3.6324	EV
	7	2.1263	3.5090	EV
	8	2.4282	3.7957	EV
	9	2.0364	3.2376	EV
	10	2.5413	3.6056	EV
	15	2.7126	3.3723	EV
	20	3.0343	3.8388	EV

Table A.176: Sign Test Comparison - EV vs. OM1 - Design 3, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	5	0.979	EV
	3	5	0.979	EV
	4	5	0.979	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	0	1.000	EV
	8	3	0.999	EV
	9	3	0.999	EV
	10	1	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	8	0.748	Same
	3	1	1.000	EV
	4	3	0.999	EV
	5	2	1.000	EV
	6	3	0.999	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	1	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.177: Paired-t Comparison - EV vs. OM1 - Design 3, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-19.1430	21.8566	Same
	3	-125.5893	50.3533	Same
	4	-127.0853	-1.9935	OM1
	5	-48.9132	2.9629	Same
	6	-53.5146	-4.8387	OM1
	7	-30.9580	6.4077	Same
	8	-73.7984	13.0887	Same
	9	-39.2816	4.0651	Same
	10	-25.9296	5.3090	Same
	15	-13.5717	12.1789	Same
	20	-6.7002	13.2530	Same
SP	2	-17.9780	6.7119	Same
	3	-9.5077	4.4612	Same
	4	-6.3089	4.2258	Same
	5	-4.5206	4.6555	Same
	6	-3.4571	4.1803	Same
	7	-3.6427	4.2968	Same
	8	-1.4210	5.4146	Same
	9	-2.2842	2.2320	Same
	10	-2.4977	2.1089	Same
	15	-2.3689	0.6817	Same
	20	-1.4238	1.2848	Same

Table A.178: Sign Test Comparison - EV vs. OM1 - Design 3, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	12	0.132	Same
	3	11	0.252	Same
	4	14	0.021	OM1
	5	13	0.058	OM1
	6	13	0.058	OM1
	7	11	0.252	Same
	8	11	0.252	Same
	9	12	0.132	Same
	10	12	0.132	Same
	15	8	0.748	Same
	20	10	0.412	Same
SP	2	10	0.412	Same
	3	10	0.412	Same
	4	10	0.412	Same
	5	10	0.412	Same
	6	9	0.588	Same
	7	10	0.412	Same
	8	8	0.748	Same
	9	9	0.588	Same
	10	9	0.588	Same
	15	11	0.252	Same
	20	8	0.748	Same

Table A.179: Paired-t Comparison - EV vs. OM1 - Design 3, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-86.3585	357.7381	Same
	3	-273.4700	351.5701	Same
	4	146.2244	435.7945	EV
	5	280.2435	454.8443	EV
	6	290.4177	538.5359	EV
	7	267.9994	426.2457	EV
	8	162.1080	493.3853	EV
	9	353.9704	539.9599	EV
	10	354.6142	546.6857	EV
	15	474.0529	583.2009	EV
	20	455.7807	584.6658	EV
SP	2	29.5744	62.1359	EV
	3	50.1225	112.3693	EV
	4	56.8869	98.5894	EV
	5	60.8278	99.1739	EV
	6	61.9814	99.2548	EV
	7	71.7026	105.0954	EV
	8	78.0026	101.2814	EV
	9	64.0738	82.0490	EV
	10	82.6953	105.4786	EV
	15	78.6372	97.8644	EV
	20	87.2860	106.7123	EV

Table A.180: Sign Test Comparison - EV vs. OM1 - Design 3, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	4	0.994	EV
	3	5	0.979	EV
	4	1	1.000	EV
	5	0	1.000	EV
	6	1	1.000	EV
	7	0	1.000	EV
	8	2	1.000	EV
	9	0	1.000	EV
	10	2	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.181: Paired-t Comparison - EV vs. OM1 - Design 4, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-1.7106	4.0182	Same
	3	2.0183	4.6021	EV
	4	0.7029	4.0607	EV
	5	2.6390	4.1825	EV
	6	0.8794	3.8101	EV
	7	1.2826	4.0041	EV
	8	0.4545	3.1679	EV
	9	1.8207	3.3220	EV
	10	0.8096	3.2816	EV
	15	0.9652	2.9014	EV
	20	2.2468	3.3191	EV
SP	2	-0.7469	1.5506	Same
	3	0.1658	1.4308	EV
	4	-0.2725	1.3219	Same
	5	0.7427	1.8517	EV
	6	1.1351	3.0401	EV
	7	1.5147	3.2111	EV
	8	1.9743	3.4189	EV
	9	1.5923	2.8184	EV
	10	2.1752	3.4088	EV
	15	2.5354	3.9006	EV
	20	2.3413	3.2956	EV

Table A.182: Sign Test Comparison - EV vs. OM1 - Design 4, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	2	1.000	EV
	4	3	0.999	EV
	5	0	1.000	EV
	6	3	0.999	EV
	7	4	0.994	EV
	8	5	0.979	EV
	9	1	1.000	EV
	10	1	1.000	EV
	15	2	1.000	EV
	20	0	1.000	EV
SP	2	12	0.132	Same
	3	4	0.994	EV
	4	9	0.588	Same
	5	3	0.999	EV
	6	2	1.000	EV
	7	2	1.000	EV
	8	0	1.000	EV
	9	1	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.183: Paired-t Comparison - EV vs. OM1 - Design 4, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-46.5832	13.7018	Same
	3	-20.2440	17.2527	Same
	4	-16.9534	11.1711	Same
	5	-10.8319	15.1724	Same
	6	-16.9231	6.3376	Same
	7	-13.0709	9.1689	Same
	8	-12.4252	8.0752	Same
	9	-21.5447	-0.2959	OM1
	10	-17.5351	0.8865	Same
	15	-13.5744	4.4047	Same
	20	-13.0227	1.4080	Same
SP	2	-19.6445	10.7692	Same
	3	-8.3955	14.3434	Same
	4	-27.4420	-2.6447	OM1
	5	-17.3668	2.7601	Same
	6	-11.6553	2.6391	Same
	7	-8.6357	4.1511	Same
	8	-11.7718	1.0539	Same
	9	-11.8918	-0.1243	OM1
	10	-7.9247	5.0989	Same
	15	-7.6509	3.5392	Same
	20	-4.1405	3.9064	Same

Table A.184: Sign Test Comparison - EV vs. OM1 - Design 4, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	10	0.412	Same
	3	10	0.412	Same
	4	11	0.252	Same
	5	9	0.588	Same
	6	8	0.748	Same
	7	12	0.132	Same
	8	9	0.588	Same
	9	14	0.021	OM1
	10	12	0.132	Same
	15	10	0.412	Same
	20	13	0.058	OM1
SP	2	9	0.588	Same
	3	7	0.868	Same
	4	14	0.021	OM1
	5	12	0.132	Same
	6	13	0.058	OM1
	7	11	0.252	Same
	8	13	0.058	OM1
	9	13	0.058	OM1
	10	12	0.132	Same
	15	10	0.412	Same
	20	10	0.412	Same

Table A.185: Paired-t Comparison - EV vs. OM1 - Design 4, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-0.7820	82.4433	Same
	3	61.2890	99.0238	EV
	4	45.6754	88.3024	EV
	5	77.5510	103.6043	EV
	6	48.4045	88.5880	EV
	7	46.3866	86.9394	EV
	8	29.2139	77.5807	EV
	9	67.9366	94.0761	EV
	10	58.0125	92.6145	EV
	15	55.8820	87.4287	EV
	20	78.1416	94.7482	EV
SP	2	39.9657	77.2973	EV
	3	54.6575	75.0407	EV
	4	59.1515	77.1980	EV
	5	72.9882	84.8715	EV
	6	71.5101	96.4639	EV
	7	74.0228	98.8804	EV
	8	78.0313	103.5767	EV
	9	74.0813	92.5331	EV
	10	80.2900	96.2202	EV
	15	89.2818	105.6970	EV
	20	84.8147	96.8205	EV

Table A.186: Sign Test Comparison - EV vs. OM1 - Design 4, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	4	0.994	EV
	3	0	1.000	EV
	4	2	1.000	EV
	5	0	1.000	EV
	6	1	1.000	EV
	7	2	1.000	EV
	8	2	1.000	EV
	9	0	1.000	EV
	10	1	1.000	EV
	15	1	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.187: Paired-t Comparison - EV vs. OM1 - Design 5, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	5.6790	11.0852	EV
	3	3.9715	9.6804	EV
	4	2.3974	9.8215	EV
	5	4.7469	11.1474	EV
	6	5.3567	9.7487	EV
	7	7.3510	11.5262	EV
	8	7.2620	9.6857	EV
	9	4.8604	10.0066	EV
	10	8.0731	10.7674	EV
	15	6.6458	10.5509	EV
	20	8.7163	10.6967	EV
SP	2	-0.0490	1.1888	Same
	3	1.2122	2.8362	EV
	4	1.4627	3.4428	EV
	5	2.2302	4.1638	EV
	6	2.3775	3.6434	EV
	7	2.4775	3.4917	EV
	8	2.5419	3.3785	EV
	9	2.7694	4.0895	EV
	10	2.6580	3.6646	EV
	15	2.8574	3.9872	EV
	20	3.4029	4.0171	EV

Table A.188: Sign Test Comparison - EV vs. OM1 - Design 5, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	1	1.000	EV
	3	1	1.000	EV
	4	2	1.000	EV
	5	1	1.000	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	0	1.000	EV
	9	2	1.000	EV
	10	0	1.000	EV
	15	1	1.000	EV
	20	0	1.000	EV
SP	2	7	0.868	Same
	3	2	1.000	EV
	4	3	0.999	EV
	5	1	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.189: Paired-t Comparison - EV vs. OM1 - Design 5, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-10.7215	43.9439	Same
	3	-26.1441	21.9533	Same
	4	-28.4098	19.2252	Same
	5	-34.4073	8.5029	Same
	6	-19.2836	8.4130	Same
	7	-11.0037	14.1569	Same
	8	-27.9220	3.7434	Same
	9	-13.0233	11.9036	Same
	10	-13.3992	5.5408	Same
	15	-20.8961	2.7932	Same
	20	-7.6520	6.7469	Same
SP	2	-28.2216	-3.6133	OM1
	3	-11.4252	2.3570	Same
	4	-7.4003	2.2893	Same
	5	1.3625	9.0093	EV
	6	-6.3530	2.4815	Same
	7	-4.7087	3.2931	Same
	8	-4.7330	2.5434	Same
	9	-3.2541	3.3129	Same
	10	-3.9464	2.6778	Same
	15	-2.4352	2.7123	Same
	20	-2.1710	2.9170	Same

Table A.190: Sign Test Comparison - EV vs. OM1 - Design 5, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	6	0.942	EV
	3	8	0.748	Same
	4	8	0.748	Same
	5	12	0.132	Same
	6	11	0.252	Same
	7	8	0.748	Same
	8	11	0.252	Same
	9	8	0.748	Same
	10	9	0.588	Same
	15	11	0.252	Same
	20	10	0.412	Same
SP	2	14	0.021	OM1
	3	11	0.252	Same
	4	14	0.021	OM1
	5	7	0.868	Same
	6	11	0.252	Same
	7	11	0.252	Same
	8	11	0.252	Same
	9	11	0.252	Same
	10	9	0.588	Same
	15	9	0.588	Same
	20	10	0.412	Same

Table A.191: Paired-t Comparison - EV vs. OM1 - Design 5, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	165.7989	260.7828	EV
	3	165.3824	255.9888	EV
	4	103.2299	254.4636	EV
	5	168.5848	270.7543	EV
	6	197.1217	259.3920	EV
	7	200.5234	271.1251	EV
	8	229.0784	261.7030	EV
	9	154.0546	268.2778	EV
	10	201.8990	256.4459	EV
	15	192.5573	252.1845	EV
	20	230.0741	269.8483	EV
SP	2	38.0611	61.7858	EV
	3	63.8900	90.2174	EV
	4	65.1619	102.7844	EV
	5	80.6512	108.1429	EV
	6	78.1784	101.3931	EV
	7	79.6672	99.4518	EV
	8	80.0730	93.5976	EV
	9	89.8150	112.9839	EV
	10	82.4709	100.8212	EV
	15	85.9409	105.2125	EV
	20	91.4060	103.9536	EV

Table A.192: Sign Test Comparison - EV vs. OM1 - Design 5, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	0	1.000	EV
	3	1	1.000	EV
	4	2	1.000	EV
	5	1	1.000	EV
	6	0	1.000	EV
	7	1	1.000	EV
	8	0	1.000	EV
	9	1	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.193: Paired-t Comparison - EV vs. OM1 - Design 6, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	2.0273	20.3197	EV
	3	12.1937	20.6467	EV
	4	9.5207	24.1614	EV
	5	17.5568	24.0282	EV
	6	15.4469	22.0531	EV
	7	15.5790	22.2444	EV
	8	14.8848	22.0336	EV
	9	18.5783	24.7895	EV
	10	19.7039	24.6226	EV
	15	20.6391	24.1916	EV
	20	21.6909	24.2083	EV
SP	2	1.3018	4.0782	EV
	3	1.7881	3.4695	EV
	4	2.5438	3.8648	EV
	5	2.5848	3.6269	EV
	6	2.7740	4.0213	EV
	7	2.8537	4.1345	EV
	8	3.4035	4.4714	EV
	9	3.1984	4.2127	EV
	10	3.0661	3.8795	EV
	15	3.1631	4.1431	EV
	20	3.2946	3.9837	EV

Table A.194: Sign Test Comparison - EV vs. OM1 - Design 6, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	0	1.000	EV
	4	2	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	1	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	5	0.979	EV
	3	1	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.195: Paired-t Comparison - EV vs. OM1 - Design 6, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-53.2920	37.8529	Same
	3	-8.8302	23.8389	Same
	4	-47.5587	35.5553	Same
	5	-7.3213	21.7958	Same
	6	-10.0226	11.7649	Same
	7	3.1920	28.6185	Same
	8	-18.8296	14.9291	Same
	9	-1.9854	24.3241	Same
	10	5.7523	21.1841	EV
	15	-5.4164	17.2124	Same
	20	8.0609	21.3128	EV
SP	2	-21.3873	-6.5206	OM1
	3	-11.3347	-3.1914	OM1
	4	-6.6919	0.5721	Same
	5	-4.6303	0.5865	Same
	6	-4.0217	0.5616	Same
	7	-3.0220	1.2778	Same
	8	-1.1626	1.5012	Same
	9	-2.4945	0.7284	Same
	10	-1.7433	1.1265	Same
	15	-1.0358	2.2236	Same
	20	-0.6263	1.9200	Same

Table A.196: Sign Test Comparison - EV vs. OM1 - Design 6, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	8	0.748	Same
	3	9	0.588	Same
	4	7	0.868	Same
	5	6	0.942	EV
	6	9	0.588	Same
	7	2	1.000	EV
	8	10	0.412	Same
	9	6	0.942	EV
	10	5	0.979	EV
	15	5	0.979	EV
	20	3	0.999	EV
SP	2	17	0.000	OM1
	3	17	0.000	OM1
	4	14	0.021	OM1
	5	13	0.058	OM1
	6	13	0.058	OM1
	7	13	0.058	OM1
	8	9	0.588	Same
	9	14	0.021	OM1
	10	7	0.868	Same
	15	12	0.132	Same
	20	10	0.412	Same

Table A.197: Paired-t Comparison - EV vs. OM1 - Design 6, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	96.2334	440.8385	EV
	3	354.0721	558.5624	EV
	4	338.2019	588.5731	EV
	5	416.8831	587.9346	EV
	6	416.4759	573.6778	EV
	7	440.7703	579.1335	EV
	8	433.9851	564.6054	EV
	9	489.8309	626.7046	EV
	10	499.0746	623.3828	EV
	15	548.1055	624.5917	EV
	20	540.0832	607.0541	EV
SP	2	55.5360	112.3372	EV
	3	62.7802	99.0277	EV
	4	75.9515	103.5087	EV
	5	78.1376	101.6731	EV
	6	77.5325	104.5101	EV
	7	83.2271	110.9537	EV
	8	90.1758	112.1568	EV
	9	88.6607	111.1849	EV
	10	85.9001	106.5394	EV
	15	86.0270	105.4424	EV
	20	87.6821	102.9528	EV

Table A.198: Sign Test Comparison - EV vs. OM1 - Design 6, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	0	1.000	EV
	4	1	1.000	EV
	5	1	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.199: Paired-t Comparison - EV vs. OM1 - Design 7, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	0.0922	4.5118	EV
	3	0.9040	4.0167	EV
	4	0.3074	3.9562	EV
	5	0.5293	4.4845	EV
	6	1.1527	4.3276	EV
	7	1.3266	3.4665	EV
	8	0.9336	3.3077	EV
	9	1.1274	3.8394	EV
	10	2.6729	3.7142	EV
	15	2.3807	3.5888	EV
	20	2.2965	3.5195	EV
SP	2	0.1828	2.6196	EV
	3	1.5870	3.0888	EV
	4	2.0021	3.6454	EV
	5	2.5210	3.8851	EV
	6	2.5654	3.8262	EV
	7	2.1886	3.6557	EV
	8	2.6667	3.8426	EV
	9	2.7132	3.9666	EV
	10	2.3174	3.2962	EV
	15	3.2337	3.8364	EV
	20	3.1557	3.8811	EV

Table A.200: Sign Test Comparison - EV vs. OM1 - Design 7, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	5	0.979	EV
	3	4	0.994	EV
	4	5	0.979	EV
	5	3	0.999	EV
	6	3	0.999	EV
	7	2	1.000	EV
	8	4	0.994	EV
	9	3	0.999	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	1	1.000	EV
SP	2	7	0.868	Same
	3	1	1.000	EV
	4	2	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.201: Paired-t Comparison - EV vs. OM1 - Design 7, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-10.6979	21.9496	Same
	3	-21.8626	9.0221	Same
	4	-8.5815	17.7468	Same
	5	-23.6361	1.8534	Same
	6	-25.4671	2.1986	Same
	7	-17.5823	5.4649	Same
	8	-12.1903	5.8634	Same
	9	-15.9691	2.8806	Same
	10	-11.6266	3.4995	Same
	15	-12.6788	0.3138	Same
	20	-3.3276	5.8426	Same
SP	2	-29.2700	1.3728	Same
	3	-21.4909	-2.7255	OM1
	4	-11.4527	3.8301	Same
	5	-7.4199	5.8835	Same
	6	-4.8163	7.7381	Same
	7	-10.7808	4.0903	Same
	8	-9.8168	-0.2221	OM1
	9	-7.6686	2.1209	Same
	10	-6.1594	4.1114	Same
	15	-5.3321	2.2076	Same
	20	-7.4339	-0.6885	OM1

Table A.202: Sign Test Comparison - EV vs. OM1 - Design 7, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	7	0.868	Same
	3	11	0.252	Same
	4	7	0.868	Same
	5	14	0.021	OM1
	6	11	0.252	Same
	7	10	0.412	Same
	8	10	0.412	Same
	9	13	0.058	OM1
	10	12	0.132	Same
	15	14	0.021	OM1
	20	8	0.748	Same
SP	2	11	0.252	Same
	3	14	0.021	OM1
	4	11	0.252	Same
	5	11	0.252	Same
	6	11	0.252	Same
	7	11	0.252	Same
	8	14	0.021	OM1
	9	15	0.006	OM1
	10	12	0.132	Same
	15	12	0.132	Same
	20	14	0.021	OM1

Table A.203: Paired-t Comparison - EV vs. OM1 - Design 7, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	23.1171	87.8225	EV
	3	41.4911	101.3028	EV
	4	13.5983	84.2567	EV
	5	57.2552	103.1730	EV
	6	56.8358	98.7463	EV
	7	62.2188	98.7250	EV
	8	47.6509	85.8831	EV
	9	70.7870	99.3372	EV
	10	82.1951	99.2936	EV
	15	74.6319	89.2743	EV
	20	71.3868	90.1656	EV
SP	2	54.8153	80.4336	EV
	3	70.0271	93.2986	EV
	4	72.8727	99.8898	EV
	5	81.5435	98.4486	EV
	6	83.4557	98.3137	EV
	7	85.1724	102.1037	EV
	8	90.3111	107.6270	EV
	9	85.5167	100.2069	EV
	10	85.5212	95.8610	EV
	15	91.8882	102.6800	EV
	20	94.7989	105.2885	EV

Table A.204: Sign Test Comparison - EV vs. OM1 - Design 7, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	1	1.000	EV
	4	4	0.994	EV
	5	2	1.000	EV
	6	1	1.000	EV
	7	2	1.000	EV
	8	2	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.205: Paired-t Comparison - EV vs. OM1 - Design 8, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-4.1607	8.8072	Same
	3	5.3406	11.0063	EV
	4	7.5989	10.2287	EV
	5	7.6337	10.2624	EV
	6	9.0036	11.1068	EV
	7	8.0440	10.6533	EV
	8	8.7421	10.7745	EV
	9	7.8262	10.2157	EV
	10	8.7084	10.8545	EV
	15	8.9330	10.9724	EV
	20	9.0705	10.4567	EV
SP	2	1.4488	4.1164	EV
	3	2.3872	3.8141	EV
	4	2.6396	3.5456	EV
	5	2.8184	3.7925	EV
	6	2.7260	3.3771	EV
	7	3.2622	3.9546	EV
	8	3.1019	3.8476	EV
	9	3.2484	4.0830	EV
	10	3.4734	4.0618	EV
	15	3.1757	3.8872	EV
	20	3.4659	4.0351	EV

Table A.206: Sign Test Comparison - EV vs. OM1 - Design 8, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	5	0.979	EV
	3	2	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	1	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.207: Paired-t Comparison - EV vs. OM1 - Design 8, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-52.1256	18.8318	Same
	3	-20.7686	13.5287	Same
	4	-30.8090	13.2973	Same
	5	-14.9426	9.1323	Same
	6	-13.3325	8.4761	Same
	7	-9.7765	8.6835	Same
	8	-18.7393	6.4059	Same
	9	-8.7093	8.0129	Same
	10	-4.2765	10.2292	Same
	15	-9.1421	1.7156	Same
	20	-4.9581	3.0586	Same
SP	2	-24.3241	-4.5174	OM1
	3	-12.1580	-1.0932	OM1
	4	-7.7608	2.4633	Same
	5	-4.9710	1.3927	Same
	6	-6.9177	-0.6318	OM1
	7	-3.1533	1.6077	Same
	8	-3.3940	3.2330	Same
	9	-2.8019	1.8154	Same
	10	-2.9005	2.5782	Same
	15	-1.3588	2.7409	Same
	20	-1.4084	2.6841	Same

Table A.208: Sign Test Comparison - EV vs. OM1 - Design 8, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	8	0.748	Same
	3	11	0.252	Same
	4	11	0.252	Same
	5	8	0.748	Same
	6	9	0.588	Same
	7	9	0.588	Same
	8	10	0.412	Same
	9	7	0.868	Same
	10	8	0.748	Same
	15	10	0.412	Same
	20	11	0.252	Same
SP	2	16	0.001	OM1
	3	14	0.021	OM1
	4	12	0.132	Same
	5	11	0.252	Same
	6	15	0.006	OM1
	7	10	0.412	Same
	8	11	0.252	Same
	9	11	0.252	Same
	10	11	0.252	Same
	15	10	0.412	Same
	20	7	0.868	Same

Table A.209: Paired-t Comparison - EV vs. OM1 - Design 8, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-46.9162	208.1845	Same
	3	164.5731	250.4065	EV
	4	225.7233	265.2253	EV
	5	224.2010	272.1496	EV
	6	237.5648	279.2732	EV
	7	216.8456	253.5093	EV
	8	242.9024	285.3054	EV
	9	220.2925	258.8332	EV
	10	226.2083	272.3399	EV
	15	245.3390	272.3137	EV
	20	238.1279	275.5154	EV
SP	2	62.5117	111.5644	EV
	3	74.3135	100.7883	EV
	4	79.5140	95.7578	EV
	5	81.3935	99.9132	EV
	6	82.3715	93.6883	EV
	7	91.0640	105.0558	EV
	8	84.5445	99.0184	EV
	9	91.0244	105.7672	EV
	10	91.9760	104.9743	EV
	15	85.5627	97.1863	EV
	20	91.9074	103.8228	EV

Table A.210: Sign Test Comparison - EV vs. OM1 - Design 8, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	4	0.994	EV
	3	1	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.211: Paired-t Comparison - EV vs. OM1 - Design 9, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	0.1449	20.6514	EV
	3	11.8912	21.3577	EV
	4	10.8228	21.7583	EV
	5	18.3210	24.1721	EV
	6	18.2418	26.1637	EV
	7	17.9993	23.8975	EV
	8	19.3680	24.3315	EV
	9	21.3799	25.0934	EV
	10	22.3602	25.1514	EV
	15	24.0512	25.8461	EV
	20	23.5877	25.8358	EV
SP	2	2.4939	4.4533	EV
	3	2.6929	4.2174	EV
	4	3.5334	4.7082	EV
	5	3.2014	4.0627	EV
	6	3.2868	4.1906	EV
	7	3.4646	4.3798	EV
	8	3.4693	4.4053	EV
	9	3.4630	4.1921	EV
	10	3.4203	4.0459	EV
	15	3.5058	3.9732	EV
	20	3.5430	4.0744	EV

Table A.212: Sign Test Comparison - EV vs. OM1 - Design 9, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	2	1.000	EV
	3	2	1.000	EV
	4	2	1.000	EV
	5	0	1.000	EV
	6	1	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.213: Paired-t Comparison - EV vs. OM1 - Design 9, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-68.9406	25.1948	Same
	3	-16.9136	18.2794	Same
	4	-23.4390	21.2322	Same
	5	2.5158	23.1687	Same
	6	-2.3056	22.9995	Same
	7	1.6732	26.2527	EV
	8	1.9579	17.0116	EV
	9	1.4808	21.1016	EV
	10	11.6439	26.5402	EV
	15	11.4010	23.7263	EV
	20	10.4061	19.8674	EV
SP	2	-10.6594	-2.4280	OM1
	3	-7.5784	-1.1186	OM1
	4	-1.9974	2.6046	Same
	5	-2.6824	0.4318	Same
	6	-1.3982	2.4040	Same
	7	0.1015	2.7829	EV
	8	-0.9616	2.9353	Same
	9	0.2736	3.2655	EV
	10	0.5286	3.0783	EV
	15	-0.5313	1.9160	Same
	20	1.3648	3.1615	EV

Table A.214: Sign Test Comparison - EV vs. OM1 - Design 9, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	8	0.748	Same
	3	9	0.588	Same
	4	7	0.868	Same
	5	7	0.868	Same
	6	4	0.994	EV
	7	4	0.994	EV
	8	5	0.979	EV
	9	7	0.868	Same
	10	3	0.999	EV
	15	1	1.000	EV
	20	1	1.000	EV
SP	2	17	0.000	OM1
	3	16	0.001	OM1
	4	10	0.412	Same
	5	12	0.132	Same
	6	10	0.412	Same
	7	7	0.868	Same
	8	8	0.748	Same
	9	8	0.748	Same
	10	6	0.942	EV
	15	10	0.412	Same
	20	1	1.000	EV

Table A.215: Paired-t Comparison - EV vs. OM1 - Design 9, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	121.1398	495.1217	EV
	3	339.6678	548.7103	EV
	4	352.1239	550.1627	EV
	5	457.3374	595.9603	EV
	6	488.7038	634.5021	EV
	7	486.6681	598.7887	EV
	8	503.4978	622.9819	EV
	9	556.3395	638.3007	EV
	10	556.4501	625.6290	EV
	15	586.8016	627.0710	EV
	20	598.3061	645.6642	EV
SP	2	73.8258	116.3000	EV
	3	75.4923	107.8230	EV
	4	89.4063	113.7801	EV
	5	86.5546	109.4907	EV
	6	85.1837	104.5791	EV
	7	89.4300	109.5426	EV
	8	91.9801	112.2935	EV
	9	89.6871	105.1090	EV
	10	88.7440	102.5682	EV
	15	92.3123	104.2611	EV
	20	90.6681	102.7681	EV

Table A.216: Sign Test Comparison - EV vs. OM1 - Design 9, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	2	1.000	EV
	3	2	1.000	EV
	4	1	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.217: Paired-t Comparison - EV vs. OM2 - Design 1, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-8.6751	2.0006	Same
	3	-4.3690	0.3794	Same
	4	-7.4455	0.9263	Same
	5	1.5335	3.4856	EV
	6	-0.4790	2.8371	Same
	7	-0.1627	2.8455	Same
	8	-0.3240	3.8647	Same
	9	-1.0232	2.5519	Same
	10	1.7106	3.7990	Same
	15	-0.5781	1.9471	Same
	20	-0.1258	2.7972	Same
SP	2	-0.8711	1.0852	Same
	3	-0.4671	1.7453	Same
	4	-0.2467	2.1567	Same
	5	-0.5364	1.0050	Same
	6	0.1705	1.6399	EV
	7	0.2364	1.6668	EV
	8	0.1986	1.9645	EV
	9	0.2841	2.0485	EV
	10	0.9501	2.0815	EV
	15	1.5864	3.4338	EV
	20	2.2004	3.4622	EV

Table A.218: Sign Test Comparison - EV vs. OM2 - Design 1, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	11	0.252	Same
	3	12	0.132	Same
	4	11	0.252	Same
	5	4	0.994	EV
	6	7	0.868	Same
	7	5	0.979	EV
	8	7	0.868	Same
	9	8	0.748	Same
	10	2	1.000	EV
	15	9	0.588	Same
	20	4	0.994	EV
SP	2	12	0.132	Same
	3	7	0.868	Same
	4	9	0.588	Same
	5	10	0.412	Same
	6	5	0.979	EV
	7	5	0.979	EV
	8	5	0.979	EV
	9	4	0.994	EV
	10	2	1.000	EV
	15	4	0.994	EV
	20	0	1.000	EV

Table A.219: Paired-t Comparison - EV vs. OM2 - Design 1, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-79.2713	4.3833	Same
	3	-54.0375	-3.1651	OM2
	4	-76.4845	7.9967	Same
	5	-44.9790	9.3405	Same
	6	-40.2683	5.5758	Same
	7	-19.5734	3.3985	Same
	8	-35.0962	-1.0838	Same
	9	-31.5678	5.4601	Same
	10	-19.5023	5.3365	Same
	15	-23.9465	-3.6505	Same
	20	-18.4229	1.2718	Same
SP	2	-10.4712	20.7918	Same
	3	-22.9025	12.5377	Same
	4	-3.1587	19.1082	Same
	5	-13.5933	11.6333	Same
	6	-19.7341	4.5653	Same
	7	-14.3004	7.5136	Same
	8	-21.3371	1.3713	Same
	9	-21.6624	3.6557	Same
	10	-18.6704	1.4873	Same
	15	-15.6052	1.2513	Same
	20	-5.7936	6.5689	Same

Table A.220: Sign Test Comparison - EV vs. OM2 - Design 1, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	12	0.132	Same
	3	14	0.021	OM2
	4	12	0.132	Same
	5	9	0.588	Same
	6	12	0.132	Same
	7	12	0.132	Same
	8	14	0.021	OM2
	9	15	0.006	OM2
	10	10	0.412	Same
	15	15	0.006	OM2
	20	14	0.021	OM2
SP	2	13	0.058	OM2
	3	8	0.748	Same
	4	5	0.979	EV
	5	7	0.868	Same
	6	11	0.252	Same
	7	12	0.132	Same
	8	12	0.132	Same
	9	13	0.058	OM2
	10	11	0.252	Same
	15	14	0.021	OM2
	20	10	0.412	Same

Table A.221: Paired-t Comparison - EV vs. OM2 - Design 1, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-159.2408	35.0014	Same
	3	-28.6133	47.2970	Same
	4	-69.9004	70.6079	Same
	5	29.0919	81.5937	EV
	6	26.8955	76.0319	EV
	7	24.8069	75.9485	EV
	8	25.6563	75.2379	EV
	9	26.1250	82.4019	EV
	10	56.1800	92.5171	EV
	15	34.8024	75.7017	EV
	20	39.3289	78.1584	EV
SP	2	22.4549	60.7528	EV
	3	37.7712	74.0375	EV
	4	58.7480	96.8100	EV
	5	50.0090	80.0921	EV
	6	62.2963	85.7204	EV
	7	60.2283	84.8086	EV
	8	63.3384	78.4352	EV
	9	63.6565	83.1023	EV
	10	75.9785	92.7090	EV
	15	77.1563	102.0171	EV
	20	81.3802	100.7429	EV

Table A.222: Sign Test Comparison - EV vs. OM2 - Design 1, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	10	0.412	Same
	3	6	0.942	EV
	4	7	0.868	Same
	5	2	1.000	EV
	6	5	0.979	EV
	7	3	0.999	EV
	8	2	1.000	EV
	9	1	1.000	EV
	10	1	1.000	EV
	15	1	1.000	EV
	20	1	1.000	EV
SP	2	1	1.000	EV
	3	1	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.223: Paired-t Comparison - EV vs. OM2 - Design 2, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-3.4419	4.2498	Same
	3	0.3688	7.3154	EV
	4	1.8943	7.5419	EV
	5	-1.7729	8.7083	Same
	6	-1.0803	7.8785	Same
	7	-5.4958	7.5697	Same
	8	-2.3315	8.2176	Same
	9	3.0646	7.9351	EV
	10	3.5494	9.3564	EV
	15	2.1704	8.4790	EV
	20	8.2541	10.6603	EV
SP	2	-0.2793	1.2943	Same
	3	-0.1921	1.2302	Same
	4	-0.3719	0.8964	Same
	5	0.3432	1.8837	EV
	6	1.2352	3.0345	EV
	7	2.0002	4.6845	EV
	8	1.3141	3.1299	EV
	9	1.4572	3.1199	EV
	10	1.9280	3.8062	EV
	15	2.4675	3.7873	EV
	20	2.5125	3.5344	EV

Table A.224: Sign Test Comparison - EV vs. OM2 - Design 2, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	8	0.748	Same
	3	6	0.942	EV
	4	4	0.994	EV
	5	2	1.000	EV
	6	5	0.979	EV
	7	5	0.979	EV
	8	4	0.994	EV
	9	1	1.000	EV
	10	3	0.999	EV
	15	3	0.999	EV
	20	0	1.000	EV
SP	2	8	0.748	Same
	3	7	0.868	Same
	4	10	0.412	Same
	5	3	0.999	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	2	1.000	EV
	9	3	0.999	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.225: Paired-t Comparison - EV vs. OM2 - Design 2, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-54.6074	12.9673	Same
	3	-25.6020	4.3374	Same
	4	-32.4951	11.7725	Same
	5	-41.8301	6.6733	Same
	6	-43.2989	3.1059	Same
	7	-54.8174	9.9754	Same
	8	-47.1762	9.3639	Same
	9	-54.2614	-2.1651	OM2
	10	-37.6813	3.4819	Same
	15	-44.6356	-12.1456	OM2
	20	-22.0342	-3.1179	OM2
SP	2	-6.4319	30.3543	Same
	3	-4.7710	22.4632	Same
	4	-6.8460	13.5406	Same
	5	-0.5958	13.1808	Same
	6	-0.8610	10.1391	Same
	7	-1.6899	10.7556	Same
	8	-0.9698	9.7138	Same
	9	-5.0554	6.4960	Same
	10	-0.3965	9.4099	Same
	15	-7.6307	-0.6073	OM2
	20	-2.3655	4.7989	Same

Table A.226: Sign Test Comparison - EV vs. OM2 - Design 2, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	13	0.058	OM2
	3	11	0.252	Same
	4	12	0.132	Same
	5	12	0.132	Same
	6	14	0.021	OM2
	7	10	0.412	Same
	8	12	0.132	Same
	9	15	0.006	OM2
	10	12	0.132	Same
	15	17	0.000	OM2
	20	15	0.006	OM2
SP	2	4	0.994	EV
	3	4	0.994	EV
	4	4	0.994	EV
	5	5	0.979	EV
	6	6	0.942	EV
	7	7	0.868	Same
	8	7	0.868	Same
	9	9	0.588	Same
	10	7	0.868	Same
	15	15	0.006	OM2
	20	7	0.868	Same

Table A.227: Paired-t Comparison - EV vs. OM2 - Design 2, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	28.6388	161.8541	EV
	3	39.8528	154.3943	EV
	4	129.9640	237.7573	EV
	5	67.6947	251.0607	EV
	6	94.0520	234.5178	EV
	7	-38.8185	183.6647	Same
	8	44.6246	247.5692	EV
	9	147.2260	229.3026	EV
	10	85.1342	225.9435	EV
	15	69.7907	234.7913	EV
	20	193.2100	260.1993	EV
SP	2	33.0161	65.4671	EV
	3	35.7283	76.6812	EV
	4	43.8272	66.4636	EV
	5	52.5947	83.5690	EV
	6	70.7345	100.0879	EV
	7	74.8917	115.6298	EV
	8	64.3982	97.8720	EV
	9	68.9908	91.7631	EV
	10	81.3853	107.0485	EV
	15	82.2812	112.1345	EV
	20	84.3848	100.5111	EV

Table A.228: Sign Test Comparison - EV vs. OM2 - Design 2, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	5	0.979	EV
	3	4	0.994	EV
	4	1	1.000	EV
	5	1	1.000	EV
	6	2	1.000	EV
	7	3	0.999	EV
	8	2	1.000	EV
	9	0	1.000	EV
	10	3	0.999	EV
	15	2	1.000	EV
	20	1	1.000	EV
SP	2	1	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.229: Paired-t Comparison - EV vs. OM2 - Design 3, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-5.0590	16.2645	Same
	3	-20.4830	14.4852	Same
	4	-0.0016	17.7006	Same
	5	9.2121	18.2627	EV
	6	8.2461	20.3226	EV
	7	9.7762	16.6971	EV
	8	0.5883	20.3186	EV
	9	11.4305	21.1615	EV
	10	13.0823	21.7798	EV
	15	18.2957	24.2431	EV
	20	18.2554	24.1296	EV
SP	2	-0.6209	0.6551	Same
	3	0.6706	2.7007	EV
	4	0.6352	2.3573	EV
	5	1.1946	3.0864	EV
	6	1.3695	3.0953	EV
	7	1.7676	3.0607	EV
	8	2.0766	3.4025	EV
	9	1.7228	2.8435	EV
	10	2.3226	3.3257	EV
	15	2.5551	3.2520	EV
	20	2.8487	3.6911	EV

Table A.230: Sign Test Comparison - EV vs. OM2 - Design 3, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	5	0.979	EV
	3	5	0.979	EV
	4	5	0.979	EV
	5	2	1.000	EV
	6	2	1.000	EV
	7	0	1.000	EV
	8	3	0.999	EV
	9	3	0.999	EV
	10	1	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	8	0.748	Same
	3	5	0.979	EV
	4	2	1.000	EV
	5	4	0.994	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	0	1.000	EV
	9	1	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.231: Paired-t Comparison - EV vs. OM2 - Design 3, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-18.3488	22.6944	Same
	3	-126.9091	49.8020	Same
	4	-128.5670	-2.2398	OM2
	5	-50.3333	4.2677	Same
	6	-51.9059	-1.9916	OM2
	7	-30.1079	6.5670	Same
	8	-72.5484	12.0343	Same
	9	-37.6057	3.2400	Same
	10	-27.9879	8.0851	Same
	15	-15.4916	8.3332	Same
	20	-7.4613	16.3784	Same
SP	2	-9.0313	29.4006	Same
	3	-3.3869	14.8224	Same
	4	-3.1000	12.1069	Same
	5	-7.0058	6.7706	Same
	6	-5.4911	7.1583	Same
	7	-3.6523	7.3210	Same
	8	-2.4533	6.1230	Same
	9	-3.3554	3.4767	Same
	10	-5.1105	2.1966	Same
	15	-3.5864	0.7388	Same
	20	-1.9528	1.5747	Same

Table A.232: Sign Test Comparison - EV vs. OM2 - Design 3, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	12	0.132	Same
	3	10	0.412	Same
	4	15	0.006	OM2
	5	11	0.252	Same
	6	13	0.058	OM2
	7	11	0.252	Same
	8	10	0.412	Same
	9	12	0.132	Same
	10	12	0.132	Same
	15	11	0.252	Same
	20	11	0.252	Same
SP	2	5	0.979	EV
	3	7	0.868	Same
	4	7	0.868	Same
	5	10	0.412	Same
	6	8	0.748	Same
	7	8	0.748	Same
	8	7	0.868	Same
	9	10	0.412	Same
	10	11	0.252	Same
	15	12	0.132	Same
	20	9	0.588	Same

Table A.233: Paired-t Comparison - EV vs. OM2 - Design 3, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-89.7287	356.8283	Same
	3	-276.4987	350.4788	Same
	4	147.1992	434.9554	EV
	5	279.6036	452.8056	EV
	6	290.5070	538.3043	EV
	7	265.6940	426.1282	EV
	8	162.5728	494.6936	EV
	9	355.9069	538.8990	EV
	10	352.3187	544.6519	EV
	15	469.9286	580.5525	EV
	20	454.8770	582.1367	EV
SP	2	20.1783	43.7228	EV
	3	41.9979	94.2874	EV
	4	50.3613	87.4377	EV
	5	54.9972	89.9732	EV
	6	57.8085	93.2423	EV
	7	67.7659	98.7802	EV
	8	74.0810	96.3265	EV
	9	61.3946	78.5646	EV
	10	79.8827	101.4860	EV
	15	77.1124	95.8898	EV
	20	85.1470	104.4884	EV

Table A.234: Sign Test Comparison - EV vs. OM2 - Design 3, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	4	0.994	EV
	3	5	0.979	EV
	4	1	1.000	EV
	5	0	1.000	EV
	6	1	1.000	EV
	7	0	1.000	EV
	8	2	1.000	EV
	9	0	1.000	EV
	10	2	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	1	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.235: Paired-t Comparison - EV vs. OM2 - Design 4, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	2.2501	6.3876	EV
	3	3.9167	7.3355	EV
	4	2.6921	5.7861	EV
	5	3.9498	6.4228	EV
	6	3.8751	6.1703	EV
	7	3.5717	5.4564	EV
	8	3.2868	6.1327	EV
	9	2.6611	4.6973	EV
	10	3.0996	5.3432	EV
	15	3.1296	5.1959	EV
	20	3.1355	4.3423	EV
SP	2	-0.8915	0.9147	Same
	3	-0.1889	0.7965	Same
	4	-0.5371	0.7492	Same
	5	0.3269	1.3446	EV
	6	0.9005	2.7765	EV
	7	1.3344	2.9879	EV
	8	1.6871	3.2508	EV
	9	1.5911	2.6944	EV
	10	1.9749	3.3546	EV
	15	2.4234	3.8667	EV
	20	2.2297	3.2132	EV

Table A.236: Sign Test Comparison - EV vs. OM2 - Design 4, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	2	1.000	EV
	3	1	1.000	EV
	4	2	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	1	1.000	EV
	8	2	1.000	EV
	9	1	1.000	EV
	10	1	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	13	0.058	OM2
	3	7	0.868	Same
	4	11	0.252	Same
	5	3	0.999	EV
	6	3	0.999	EV
	7	2	1.000	EV
	8	0	1.000	EV
	9	1	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.237: Paired-t Comparison - EV vs. OM2 - Design 4, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-46.2974	13.5726	Same
	3	-18.4561	17.9577	Same
	4	-16.0613	12.0114	Same
	5	-10.1493	14.7255	Same
	6	-15.5454	7.5254	Same
	7	-15.4546	7.1360	Same
	8	-11.2519	8.2256	Same
	9	-21.2640	0.4110	Same
	10	-16.5648	1.4940	Same
	15	-13.1528	4.6459	Same
	20	-11.1571	2.6073	Same
SP	2	-6.4791	24.0198	Same
	3	-1.7645	21.0244	Same
	4	-23.1257	3.5038	Same
	5	-15.7062	4.2401	Same
	6	-8.0201	6.6030	Same
	7	-4.1452	8.4839	Same
	8	-10.1625	2.6574	Same
	9	-12.7543	-0.3612	OM2
	10	-8.3674	5.1738	Same
	15	-7.3255	4.7405	Same
	20	-3.6707	4.1976	Same

Table A.238: Sign Test Comparison - EV vs. OM2 - Design 4, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	11	0.252	Same
	3	10	0.412	Same
	4	11	0.252	Same
	5	8	0.748	Same
	6	8	0.748	Same
	7	9	0.588	Same
	8	9	0.588	Same
	9	14	0.021	OM2
	10	13	0.058	OM2
	15	7	0.868	Same
	20	11	0.252	Same
SP	2	7	0.868	Same
	3	8	0.748	Same
	4	11	0.252	Same
	5	10	0.412	Same
	6	9	0.588	Same
	7	10	0.412	Same
	8	11	0.252	Same
	9	14	0.021	OM2
	10	13	0.058	OM2
	15	10	0.412	Same
	20	10	0.412	Same

Table A.239: Paired-t Comparison - EV vs. OM2 - Design 4, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-0.4558	82.4848	Same
	3	61.0562	98.6070	EV
	4	47.1948	89.7044	EV
	5	77.4647	103.6320	EV
	6	46.8092	87.4923	EV
	7	48.9760	87.5262	EV
	8	28.8510	75.7379	EV
	9	67.8790	93.9904	EV
	10	59.7310	94.1698	EV
	15	56.4187	89.5333	EV
	20	79.1070	95.3624	EV
SP	2	30.6475	61.0595	EV
	3	47.1193	65.0435	EV
	4	55.7472	73.7371	EV
	5	68.5364	80.6304	EV
	6	67.8578	91.5554	EV
	7	70.8722	94.2661	EV
	8	73.4998	96.0910	EV
	9	71.5772	91.3340	EV
	10	78.0562	92.2871	EV
	15	88.1878	105.8610	EV
	20	83.5268	94.2169	EV

Table A.240: Sign Test Comparison - EV vs. OM2 - Design 4, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	4	0.994	EV
	3	0	1.000	EV
	4	2	1.000	EV
	5	0	1.000	EV
	6	1	1.000	EV
	7	2	1.000	EV
	8	2	1.000	EV
	9	0	1.000	EV
	10	1	1.000	EV
	15	1	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.241: Paired-t Comparison - EV vs. OM2 - Design 5, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	5.7677	11.1599	EV
	3	3.9484	9.7354	EV
	4	2.3717	9.8142	EV
	5	4.6906	11.2922	EV
	6	5.4611	9.7127	EV
	7	7.2454	11.5343	EV
	8	7.1818	9.6681	EV
	9	5.1603	9.9853	EV
	10	8.1545	10.8302	EV
	15	6.6463	10.5688	EV
	20	8.7735	10.7136	EV
SP	2	-1.3451	0.0429	Same
	3	0.3154	1.4866	EV
	4	0.6752	2.3768	EV
	5	1.5848	3.3362	EV
	6	1.7671	3.0381	EV
	7	1.9902	2.9518	EV
	8	2.0610	2.8443	EV
	9	2.4828	3.7643	EV
	10	2.3387	3.2622	EV
	15	2.6389	3.7511	EV
	20	3.1786	3.7962	EV

Table A.242: Sign Test Comparison - EV vs. OM2 - Design 5, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	1	1.000	EV
	3	1	1.000	EV
	4	2	1.000	EV
	5	1	1.000	EV
	6	2	1.000	EV
	7	1	1.000	EV
	8	0	1.000	EV
	9	2	1.000	EV
	10	0	1.000	EV
	15	1	1.000	EV
	20	0	1.000	EV
SP	2	13	0.058	OM2
	3	5	0.979	EV
	4	5	0.979	EV
	5	1	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.243: Paired-t Comparison - EV vs. OM2 - Design 5, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-9.7231	45.3802	Same
	3	-24.8579	21.8490	Same
	4	-27.3348	17.8609	Same
	5	-34.4609	7.6830	Same
	6	-18.8554	9.2407	Same
	7	-10.1097	15.1813	Same
	8	-27.8040	3.7066	Same
	9	-13.8900	11.2567	Same
	10	-14.1585	5.0842	Same
	15	-18.3793	4.2384	Same
	20	-6.9571	8.6255	Same
SP	2	-26.4955	20.7299	Same
	3	-9.2874	15.4777	Same
	4	-3.1366	9.4513	Same
	5	0.2536	13.9505	EV
	6	-10.1460	4.0089	Same
	7	-5.7419	4.4054	Same
	8	-6.2182	2.0132	Same
	9	-3.8350	4.5240	Same
	10	-5.4502	2.7360	Same
	15	-3.5562	2.0247	Same
	20	-2.8376	1.4688	Same

Table A.244: Sign Test Comparison - EV vs. OM2 - Design 5, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	6	0.942	EV
	3	8	0.748	Same
	4	8	0.748	Same
	5	12	0.132	Same
	6	8	0.748	Same
	7	6	0.942	EV
	8	12	0.132	Same
	9	8	0.748	Same
	10	12	0.132	Same
	15	10	0.412	Same
	20	8	0.748	Same
SP	2	9	0.588	Same
	3	7	0.868	Same
	4	8	0.748	Same
	5	6	0.942	EV
	6	12	0.132	Same
	7	12	0.132	Same
	8	11	0.252	Same
	9	12	0.132	Same
	10	13	0.058	OM2
	15	9	0.588	Same
	20	11	0.252	Same

Table A.245: Paired-t Comparison - EV vs. OM2 - Design 5, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	165.7228	261.3846	EV
	3	165.7336	254.4903	EV
	4	105.7766	256.6250	EV
	5	167.2434	271.0341	EV
	6	196.0540	258.9346	EV
	7	200.2157	271.4411	EV
	8	228.4449	261.8413	EV
	9	152.8864	267.5062	EV
	10	201.1504	256.8212	EV
	15	189.4806	252.0514	EV
	20	230.5717	271.4953	EV
SP	2	25.0102	41.9903	EV
	3	52.0236	73.8470	EV
	4	57.0099	90.3971	EV
	5	72.8220	97.7527	EV
	6	71.7388	92.8701	EV
	7	74.2969	93.6541	EV
	8	75.2291	87.7645	EV
	9	84.9039	107.6507	EV
	10	78.2103	96.2580	EV
	15	84.5549	104.0385	EV
	20	89.9452	103.2430	EV

Table A.246: Sign Test Comparison - EV vs. OM2 - Design 5, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	0	1.000	EV
	3	1	1.000	EV
	4	2	1.000	EV
	5	1	1.000	EV
	6	0	1.000	EV
	7	1	1.000	EV
	8	0	1.000	EV
	9	1	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.247: Paired-t Comparison - EV vs. OM2 - Design 6, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	1.9117	20.2687	EV
	3	12.1750	20.7057	EV
	4	9.5994	24.1847	EV
	5	17.7605	24.1482	EV
	6	15.2782	21.8066	EV
	7	15.6656	22.3190	EV
	8	14.8984	22.0322	EV
	9	18.4806	24.8428	EV
	10	19.8387	24.6426	EV
	15	20.6580	24.2569	EV
	20	21.5976	24.0492	EV
SP	2	-1.3844	0.3518	Same
	3	0.4768	1.5287	EV
	4	1.2840	2.3648	EV
	5	1.6815	2.6061	EV
	6	1.9993	3.1501	EV
	7	2.2041	3.3612	EV
	8	2.7968	3.8087	EV
	9	2.7201	3.6464	EV
	10	2.6455	3.4224	EV
	15	2.9101	3.8590	EV
	20	3.0852	3.8096	EV

Table A.248: Sign Test Comparison - EV vs. OM2 - Design 6, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	0	1.000	EV
	4	2	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	1	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	11	0.252	Same
	3	5	0.979	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.249: Paired-t Comparison - EV vs. OM2 - Design 6, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-54.1910	37.0410	Same
	3	-8.5717	24.6076	Same
	4	-47.9670	33.6135	Same
	5	-8.6510	21.0575	Same
	6	-10.1469	12.8353	Same
	7	3.9159	30.1664	EV
	8	-18.5501	17.0272	Same
	9	-1.1229	26.0792	Same
	10	6.1607	22.7111	EV
	15	-7.4835	16.2570	Same
	20	6.8445	21.8947	EV
SP	2	-44.2127	-7.5687	OM2
	3	-27.4292	-2.5387	OM2
	4	-13.9045	2.2892	Same
	5	-14.3410	1.6969	Same
	6	-11.0404	1.1704	Same
	7	-8.4385	1.7248	Same
	8	-8.2570	-0.6692	OM2
	9	-6.1648	0.8354	Same
	10	-6.4303	0.0124	Same
	15	-1.7130	1.0068	Same
	20	-3.1180	0.6716	Same

Table A.250: Sign Test Comparison - EV vs. OM2 - Design 6, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	9	0.588	Same
	3	8	0.748	Same
	4	7	0.868	Same
	5	8	0.748	Same
	6	8	0.748	Same
	7	1	1.000	EV
	8	10	0.412	Same
	9	6	0.942	EV
	10	5	0.979	EV
	15	7	0.868	Same
	20	3	0.999	EV
SP	2	14	0.021	OM2
	3	15	0.006	OM2
	4	12	0.132	Same
	5	12	0.132	Same
	6	12	0.132	Same
	7	13	0.058	OM2
	8	15	0.006	OM2
	9	12	0.132	Same
	10	14	0.021	OM2
	15	11	0.252	Same
	20	11	0.252	Same

Table A.251: Paired-t Comparison - EV vs. OM2 - Design 6, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	96.7858	440.5401	EV
	3	353.4212	558.0768	EV
	4	337.4082	590.3425	EV
	5	418.4285	588.5744	EV
	6	417.1367	576.2066	EV
	7	439.9828	578.6659	EV
	8	437.6056	567.8148	EV
	9	490.0334	626.0593	EV
	10	500.2742	625.1865	EV
	15	545.2439	622.5022	EV
	20	539.0287	606.4549	EV
SP	2	32.9882	69.2990	EV
	3	47.6581	75.4684	EV
	4	63.2652	86.4206	EV
	5	68.1981	88.7395	EV
	6	69.1403	93.8393	EV
	7	75.9867	101.3692	EV
	8	83.0907	103.8464	EV
	9	82.8422	104.4227	EV
	10	80.3612	99.7571	EV
	15	82.2852	101.0313	EV
	20	85.1490	99.5380	EV

Table A.252: Sign Test Comparison - EV vs. OM2 - Design 6, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	0	1.000	EV
	4	1	1.000	EV
	5	1	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.253: Paired-t Comparison - EV vs. OM2 - Design 7, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	1.3703	5.5249	EV
	3	1.4278	4.4597	EV
	4	0.6165	4.3131	EV
	5	0.8968	4.7219	EV
	6	1.4168	4.5938	EV
	7	1.5136	3.6772	EV
	8	1.1675	3.4895	EV
	9	1.1518	3.9220	EV
	10	2.7701	3.8233	EV
	15	2.6112	3.6959	EV
	20	2.3183	3.5560	EV
SP	2	-1.4442	0.0703	Same
	3	0.1380	1.4207	EV
	4	0.9924	2.4143	EV
	5	1.6512	2.8619	EV
	6	1.8512	3.0207	EV
	7	1.6588	3.0230	EV
	8	2.1123	3.3584	EV
	9	2.2447	3.4214	EV
	10	1.9501	2.9396	EV
	15	3.0367	3.5698	EV
	20	2.9686	3.7655	EV

Table A.254: Sign Test Comparison - EV vs. OM2 - Design 7, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	4	0.994	EV
	4	4	0.994	EV
	5	2	1.000	EV
	6	3	0.999	EV
	7	2	1.000	EV
	8	4	0.994	EV
	9	3	0.999	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	1	1.000	EV
SP	2	10	0.412	Same
	3	7	0.868	Same
	4	2	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	1	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.255: Paired-t Comparison - EV vs. OM2 - Design 7, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-1.5154	33.4325	Same
	3	-18.5239	13.2623	Same
	4	-6.7772	22.1405	Same
	5	-21.6241	4.3747	Same
	6	-23.8505	4.4929	Same
	7	-17.6415	8.2144	Same
	8	-9.5621	8.2727	Same
	9	-15.8021	4.7145	Same
	10	-11.0781	5.8900	Same
	15	-10.2158	2.3892	Same
	20	-2.8709	6.4770	Same
SP	2	-42.7810	7.5269	Same
	3	-18.8670	5.2578	Same
	4	-9.0890	7.1248	Same
	5	-12.0513	5.8435	Same
	6	-6.5375	5.2872	Same
	7	-10.9870	4.6267	Same
	8	-8.7039	0.8007	Same
	9	-8.2184	-0.5450	OM2
	10	-7.2507	3.3048	Same
	15	-6.1714	2.1854	Same
	20	-8.1257	-1.0882	OM2

Table A.256: Sign Test Comparison - EV vs. OM2 - Design 7, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	6	0.942	EV
	3	11	0.252	Same
	4	7	0.868	Same
	5	12	0.132	Same
	6	11	0.252	Same
	7	10	0.412	Same
	8	9	0.588	Same
	9	11	0.252	Same
	10	10	0.412	Same
	15	13	0.058	OM2
	20	6	0.942	EV
SP	2	10	0.412	Same
	3	11	0.252	Same
	4	7	0.868	Same
	5	9	0.588	Same
	6	11	0.252	Same
	7	9	0.588	Same
	8	12	0.132	Same
	9	13	0.058	OM2
	10	11	0.252	Same
	15	12	0.132	Same
	20	15	0.006	OM2

Table A.257: Paired-t Comparison - EV vs. OM2 - Design 7, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	44.3465	103.5424	EV
	3	49.3047	107.9290	EV
	4	23.0899	87.3628	EV
	5	61.9402	106.4000	EV
	6	61.1446	102.9056	EV
	7	64.1145	102.4513	EV
	8	50.5379	88.6180	EV
	9	72.6542	100.3769	EV
	10	83.0035	101.8591	EV
	15	75.9602	91.8741	EV
	20	72.7214	90.4177	EV
SP	2	33.7805	49.4836	EV
	3	54.9523	74.1642	EV
	4	62.0340	85.5389	EV
	5	72.2877	89.1088	EV
	6	75.7420	90.2627	EV
	7	79.0730	94.1450	EV
	8	83.3146	99.7183	EV
	9	80.4973	94.6545	EV
	10	80.7633	90.8742	EV
	15	89.2893	100.6285	EV
	20	91.9070	102.8269	EV

Table A.258: Sign Test Comparison - EV vs. OM2 - Design 7, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	2	1.000	EV
	3	1	1.000	EV
	4	4	0.994	EV
	5	1	1.000	EV
	6	1	1.000	EV
	7	2	1.000	EV
	8	2	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.259: Paired-t Comparison - EV vs. OM2 - Design 8, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-2.2702	9.8225	Same
	3	6.0689	11.6607	EV
	4	8.1156	10.6848	EV
	5	7.9579	10.5781	EV
	6	9.2735	11.3842	EV
	7	8.1806	10.7481	EV
	8	8.9239	10.9947	EV
	9	8.0637	10.5129	EV
	10	8.8873	11.0648	EV
	15	8.9875	11.0107	EV
	20	9.0602	10.4130	EV
SP	2	-1.4269	0.2616	Same
	3	0.3857	1.5470	EV
	4	1.3034	2.0458	EV
	5	1.8072	2.6018	EV
	6	1.9025	2.4434	EV
	7	2.5065	3.1376	EV
	8	2.5140	3.1486	EV
	9	2.6426	3.4103	EV
	10	2.9723	3.5016	EV
	15	2.8757	3.5848	EV
	20	3.2742	3.8838	EV

Table A.260: Sign Test Comparison - EV vs. OM2 - Design 8, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	5	0.979	EV
	3	2	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	14	0.021	OM2
	3	5	0.979	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.261: Paired-t Comparison - EV vs. OM2 - Design 8, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-39.6418	29.2058	Same
	3	-13.2727	21.2172	Same
	4	-27.6016	17.2951	Same
	5	-12.8071	12.2101	Same
	6	-10.4815	11.8769	Same
	7	-7.2312	10.2331	Same
	8	-17.4101	8.3501	Same
	9	-8.1379	8.6491	Same
	10	-3.4160	11.1668	Same
	15	-9.1260	2.7567	Same
	20	-5.2680	3.3657	Same
SP	2	-42.3957	-12.8290	OM2
	3	-31.8310	-3.1373	OM2
	4	-17.6046	-1.5884	OM2
	5	-11.8671	1.0203	Same
	6	-15.5835	-2.7694	OM2
	7	-10.9015	-1.4889	OM2
	8	-7.4791	0.0294	Same
	9	-6.9479	0.1651	Same
	10	-6.5303	-0.3639	OM2
	15	-4.5706	1.1109	Same
	20	-3.1587	0.8817	Same

Table A.262: Sign Test Comparison - EV vs. OM2 - Design 8, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	7	0.868	Same
	3	6	0.942	EV
	4	8	0.748	Same
	5	6	0.942	EV
	6	9	0.588	Same
	7	8	0.748	Same
	8	8	0.748	Same
	9	7	0.868	Same
	10	7	0.868	Same
	15	10	0.412	Same
	20	11	0.252	Same
SP	2	16	0.001	OM2
	3	13	0.058	OM2
	4	15	0.006	OM2
	5	13	0.058	OM2
	6	14	0.021	OM2
	7	15	0.006	OM2
	8	12	0.132	Same
	9	15	0.006	OM2
	10	13	0.058	OM2
	15	12	0.132	Same
	20	12	0.132	Same

Table A.263: Paired-t Comparison - EV vs. OM2 - Design 8, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-2.4078	234.6349	Same
	3	176.4491	260.9307	EV
	4	232.2032	272.4497	EV
	5	231.1237	278.0202	EV
	6	240.3105	281.6876	EV
	7	220.4243	257.2052	EV
	8	243.6004	287.4151	EV
	9	222.2219	260.2743	EV
	10	228.0442	274.8922	EV
	15	247.0854	273.6711	EV
	20	238.9056	277.8431	EV
SP	2	35.9196	66.2037	EV
	3	55.3714	76.2995	EV
	4	65.4989	78.9159	EV
	5	69.9678	85.8850	EV
	6	72.1565	82.4851	EV
	7	82.2147	95.9248	EV
	8	76.7362	90.3134	EV
	9	84.3245	98.2534	EV
	10	85.8788	97.8249	EV
	15	81.8905	93.1185	EV
	20	89.0427	100.5834	EV

Table A.264: Sign Test Comparison - EV vs. OM2 - Design 8, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	3	0.999	EV
	3	1	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.265: Paired-t Comparison - EV vs. OM2 - Design 9, Distribution 1

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	3.4936	23.5546	EV
	3	12.9664	22.2870	EV
	4	11.5840	22.3653	EV
	5	18.7920	24.5901	EV
	6	18.5853	26.3559	EV
	7	18.2487	24.1325	EV
	8	19.6605	24.6052	EV
	9	21.6017	25.2851	EV
	10	22.6578	25.3259	EV
	15	24.1193	25.9746	EV
	20	23.6744	25.9586	EV
SP	2	-1.6125	-0.5790	OM2
	3	0.3936	1.5489	EV
	4	1.8118	2.8279	EV
	5	1.9538	2.7083	EV
	6	2.2581	3.0942	EV
	7	2.5832	3.4164	EV
	8	2.7082	3.5658	EV
	9	2.8236	3.4804	EV
	10	2.8550	3.4376	EV
	15	3.1193	3.6082	EV
	20	3.2361	3.7843	EV

Table A.266: Sign Test Comparison - EV vs. OM2 - Design 9, Distribution 1

Error	Replications	Count	p-value	Conclusion
WP	2	2	1.000	EV
	3	2	1.000	EV
	4	2	1.000	EV
	5	0	1.000	EV
	6	1	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	18	0.000	OM2
	3	5	0.979	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Table A.267: Paired-t Comparison - EV vs. OM2 - Design 9, Distribution 5

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	-64.8835	42.5646	Same
	3	-7.3784	26.0522	Same
	4	-18.0763	27.0721	Same
	5	6.5622	26.5557	EV
	6	0.6673	26.4362	EV
	7	4.3461	27.4507	EV
	8	3.9738	19.5121	EV
	9	3.6102	21.4931	EV
	10	11.6916	27.4298	EV
	15	13.2377	25.0281	EV
	20	11.0983	20.9674	EV
SP	2	-72.5056	-40.7397	OM2
	3	-36.5611	-17.2602	OM2
	4	-24.5334	-9.1674	OM2
	5	-17.9777	-6.5923	OM2
	6	-13.9795	-5.7492	OM2
	7	-9.3472	-2.4714	OM2
	8	-8.5112	-1.6871	OM2
	9	-5.0783	0.6473	Same
	10	-6.5166	-1.1333	OM2
	15	-3.7868	0.0320	Same
	20	-1.8811	0.1081	Same

Table A.268: Sign Test Comparison - EV vs. OM2 - Design 9, Distribution 5

Error	Replications	Count	p-value	Conclusion
WP	2	6	0.942	EV
	3	5	0.979	EV
	4	7	0.868	Same
	5	6	0.942	EV
	6	4	0.994	EV
	7	4	0.994	EV
	8	4	0.994	EV
	9	5	0.979	EV
	10	2	1.000	EV
	15	1	1.000	EV
	20	2	1.000	EV
SP	2	18	0.000	OM2
	3	19	0.000	OM2
	4	16	0.001	OM2
	5	16	0.001	OM2
	6	16	0.001	OM2
	7	16	0.001	OM2
	8	14	0.021	OM2
	9	13	0.058	OM2
	10	15	0.006	OM2
	15	12	0.132	EV
	20	15	0.006	OM2

Table A.269: Paired-t Comparison - EV vs. OM2 - Design 9, Distribution 13

Error	Replications	Paired-t Confidence Interval		Conclusion
WP	2	179.6890	548.7617	EV
	3	362.8529	563.6973	EV
	4	365.6821	560.1141	EV
	5	463.8380	602.9592	EV
	6	494.0403	637.7421	EV
	7	491.0046	602.8586	EV
	8	507.1832	627.0178	EV
	9	558.5094	640.8128	EV
	10	560.2092	627.4725	EV
	15	588.2841	627.4296	EV
	20	598.8936	647.4131	EV
SP	2	39.5910	64.2984	EV
	3	53.8743	77.8019	EV
	4	71.0658	91.5207	EV
	5	72.6022	92.3253	EV
	6	73.9302	90.8603	EV
	7	79.6718	97.4472	EV
	8	83.2521	101.2781	EV
	9	81.7767	96.1727	EV
	10	81.4186	94.1384	EV
	15	87.5570	98.8346	EV
	20	87.1901	98.8522	EV

Table A.270: Sign Test Comparison - EV vs. OM2 - Design 9, Distribution 13

Error	Replications	Count	p-value	Conclusion
WP	2	1	1.000	EV
	3	1	1.000	EV
	4	1	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV
SP	2	0	1.000	EV
	3	0	1.000	EV
	4	0	1.000	EV
	5	0	1.000	EV
	6	0	1.000	EV
	7	0	1.000	EV
	8	0	1.000	EV
	9	0	1.000	EV
	10	0	1.000	EV
	15	0	1.000	EV
	20	0	1.000	EV

Appendix B. Blue Dart

Test and Evaluation (T&E) is a crucial part of the Defense Acquisition Management System. T&E needs to provide accurate and relevant assessments of system performance and provide early identification of deficiencies which allow for corrective actions to take place. However, limited budgets impact the amount of test that can occur. The ability for T&E to provide statistical assertions is greatly impacted by any forced reduction in the T&E effort. Experimental design methods seek to improve the efficiency and effectiveness of TE in austere budgetary environments.

Design of Experiments (DOE) is a systematic methodology to plan, conduct, and analyze an experiment in a more efficient and effective manner by maximizing the insights gained in system performance for the effort expended in experimental, or test, resources expended. The DoD has all but mandated the use of DOE throughout the acquisition developmental and operational life cycle.

DOE is not, however, without limitations, especially when few experimental replications are used, which is often the case in Air Force T&E. DOE is often limited with experimental runs cannot be accomplished in the ideal, randomized fashion, a situation known as restricted randomization. A split-plot experimental design is used, and analyzed, when the restricted randomization situation arises.

Bibliography

1. Andrews, Donald W. K. and Sukjin Han. "Invalidity of the Bootstrap and the m out of n bootstrap for interval endpoints defined by moment inequalities". *Cowles Foundation Discussion Paper*, 1671, July 2008.

2. Bickel, Peter J. and Anat Sakov. "On the choice of m in the m out of n bootstrap and confidence bounds for extrema". *Statistica Sinica*, 18:967–985, 2008.

3. Bisgaard, Søren and André L.S.de Pinho. "The Error Structure of Split-Plot Experiments". *Quality Engineering*, 16(4):671–675, 2004.

4. Booth, James G., Peter Hall, and Andrew T. A. Wood. "Balanced Importance Resampling for the Bootstrap". *The Annals of Statistics*, 21(1):286–298, 1993.

5. Box, George. *Must We Randomize Our Experiment?* Technical Report 47, Center for Quality and Productivity Improvement, University of Wisconsin-Madison, Madison, Wisconsin, December 1989.

6. Cohen, Captain Allen. *Examining Split-Plot Designs for Developmental and Operational Testing*. Master's thesis, Air Force Institute of Technology, March 2009.

7. Cohen, Michael L., John E. Rolph, , and Duane L. Steffey. *Statistics, Testing, and Defense Acquisition: New Approaches and Methodological Improvements*. National Academy Press, Washington D.C., 1998. ISBN 0-309-06551-8.

8. Corporation, Science Applications International. *Best Practices Applicable to DOD Developmental Test and Evaluation*. Technical report, Directorate of Test, System Engineering and Evaluation, Washington D.C., June 1999.

9. Davidson, James, Andrea Monticini, and David Peel. "Implementing the wild bootstrap using a two-point distribution". *Economics Letters*, 96:309–315, 2007.

10. Davidson, Russel and Emmanuel Flachaire. "The Wild Bootstrap, Tamed at Last". *Queen's Economics Department Working Paper*, 1000, 2001.

11. Davison, A.C. and D.V. Hinkley. *Bootstrap Methods and their Applications*. Cambridge University Press, Cambridge, United Kingdom, 1997. ISBN 0-521-57391-2.

12. DiCiccio, THomas J. and Bradley Efron. "More accurate confidence intervals in exponential families". *Biometrika*, 79:231–245, 1992.

13. DiCiccio, Thomas J. and Bradley Efron. "Bootstrap Confidence Intervals". *Statistical Science*, 11(3):189–228, 1996.

14. Efron, Bradley. "Bootstrap Methods: Another Look at the Jackknife". *The Annals of Statistics*, 7(1):1–26, 1979.

15. Efron, Bradley. "Better Bootstrap Confidence Intervals". *Journal of the American Statistical Association*, 82(397):171–200, 1987.

16. Efron, Bradley and Gail Gong. "A leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation". *The American Statistician*, 37(1):36–48, 1983.

17. Efron, Bradley and C. Stein. "The Jackknife Estimate of Variance". *The Annals of Statistics*, 9(3):586–596, 1981.

18. Efron, Bradley and Robert J. Tibshirani. "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy". *Statistical Science*, 1(1):54–77, 1986.

19. Efron, Bradley and Robert J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, NY, 1993. ISBN 0-412-04231-2.

20. Good, Phillip. *Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypothesis*. Springer, 2 edition, 2000.

21. Hall, Peter. "Theoretical Comparison of Bootstrap Confidence Intervals". *The Annals of Statistics*, 16(3):927–953, 1988.

22. Hall, Peter, Thomas J. DiCicio, and Joseph P. Romano. "On Smoothing and the Bootstrap". *The Annals of Statistics*, 17(2):692–704, 1989.

23. Horvath, Daniel, Rafiq Noorani, and Mel Mendelson. "Improvement of Surface Roughness on ABS 400 Polymer Using Design of Experiments (DOE)". *Materials Science Forum*, 561-565:2389–2392, 2007.

24. Kowalski, Scott M. and Kevin Potcner. "How to Analyze a Split-Plot Experiment". *Quality Progress*, 37(12):67–74, December 2004.

25. Kuehl, Robert O. *Design of Experiments: Statistical Principles of Research Design and Analysis*. Brooks/Cole, 2 edition, 2000. ISBN 978-0-534-36834-0.

26. Lahiri, S. N. "Theoretical Comparisons of Block Bootstrap Methods". *The Annals of Statistics*, 27(1):386–404, 1999.

27. Lee, Stephen M. S. and G. Alastair Young. "Asymptotic iterated bootstrap confidence intervals". *The Annals of Statistics*, 23(4):1301–1330, 1995.

28. Liu, Alex Zhi, Randy A. Weidman, Saujanya Gosangari, Christina D. Mayberry, John Trolinger, Haresh Surati, Victor Munoz, Qi Fang, and Shingai Majuru. "Implementing DOE for Product Develoment of a Generic Softgel Capsule (SGC) Product". *American Association of Pharmaceutical Scientists*, 11(S2), 2009.

29. Lorenzen, Thomas J. "Randomization and blocking in the design of experiments". *Communications in Statistics - Theory and Methods*, 13(21):2601–2623, 1984.

30. Mammen, Enno. "Bootstrap and Wild Bootstrap for High Dimensional Linear Models". *The Annals of Statistics*, 21(1):255–285, 1993.

31. Miller, Rupert G. "A trustworthy Jackknife". *The Annals of Mathematical Statistics*, 35(4):1594–1605, 1964.

32. Miller, Rupert G. "The Jackknife – A Review". *Biometrika*, 61(1):1–15, 1974.

33. Montgomery, Douglas C. *Design and Analysis of Experiments.* John Wiley & Sons, Inc., New York, NY, 7 edition, 2009. ISBN 978-0-470-12866-4.

34. Moreb, Ahmad A. and Mehmet Savsar. "Textile quality control using design of experiments". *Proceedings of the 8th WSEAS Conference on Mathematics and Computers in Business and Economics*, volume 8, 183–188. World Scientific and Engineering Academy and Society, Vancouver, Canada, 2007.

35. Nankervis, John C. "Computational algorithms for double bootstrao confidence intervals". *Computational Statistics and Data Analysis*, 49:461–475, 2005.

36. Nigam, A. K. and J. N. K. Rao. "On Balanced Bootstrap for Statified Multistage Samples". *Statistica Sinica*, 6:199–214, 1996.

37. Odén, Anders and Hans Wedel. "Arguments for Fisher's Permutation Test". *The Annals of Statistics*, 3(2):518–520, 1975.

38. Pwasong, A. D. and D. N. Choji. "The application of the Split-Plot Design in the Analysis of Experimental Rabbit Feeds Data". *Science World Journal*, 3(2):73–77, 2008.

39. Rangaswamy, R. *A Textbook of Agricultural Statistics.* New Age International Publishers, Daryaganj, New Delhi, 1995. ISBN 81-224-0758-7.

40. Rudner, Lawrence M. and Mary Morello Shafer. "Resampling: A Marriage of Computers and Statistics". *Practical Assessment, Research & Evaluation*, 3(5), 1992.

41. Silverman, B.W. and G.A. Young. "The bootstrap: To smooth or not to smooth?" *Biometrika*, 74:469–479, 1987.

42. Starkey, Michael, Jeff Aughton, and Roger Brewin. "Extending Process Thinking: Design of Experiments in Sales and Marketing". *The TQM Magazine*, 9(6):434–439, 1997.

43. Stine, Robert. "An Introduction to Bootstrap Methods". *Sociological Methods and Research*, 18(2/3):243–291, 1989/1990.

44. Stone, M. "Cross-Validation: a review". *Statistics*, 9(1):127–139, 1978.

45. Tucker, Major Aaron A., Gregory T. Hutto, and Cihan H. Dagli. "Application of Design of Experiments to Flight Test: A Case Study". United States Air Force, February 2008.

46. Uraibi, Hassan S., Habshah Midi, Bashar A. Talib, and Jabar H. Yousif. "Linear Regression Model Selection Based on Robust Bootstrapping Techniques". *American Journal of Applied Sciences*, 6(6):1191–1198, 2009.

47. USD(AT&L). *Operation of the Defense Acquisition System.* Department of Defense, Washington D.C., 5000.02 edition, December 2008.

48. Vinson, Donald E. and William J. Lundstrom. “The Use of Students as Experimental Subjects in Marketing Research”. *Journal of the Academy of Marketing Science*, 6(1-2):114–125, December 1978.

49. Vuchov, Ivan N. and Lidia N. Boyadijieva. *Quality Improvement with Design of Experiments*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 2001. ISBN 0-7923-6827-4.

50. Wehrens, Ron, Hein Putter, and Lutgarde M. C. Buydens. “The bootstrap: a tutorial”. *Chemometrics and Intelligent Laboratory Systems*, 54:35–52, 2000.

51. Young, G. Alastair. “Bootstrap: More than a Stab in the Dark?” *Statistical Science*, 9(3):382–415, 1994.

52. Yu, Chong Ho. “Resampling methods: concepts, applications, and justification”. *Practical Assessment, Research & Evaluation*, 8(19), 2003.

REPORT DOCUMENTATION PAGE			Form Approved OMB No. 074-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From – To)*
25-03-2010	**Graduate Research Project**	Sept 2008 – Mar 2010

4. TITLE AND SUBTITLE	**5a. CONTRACT NUMBER**
An Empirical Study of Re-sampling Techniques as a Method for Improving Error Estimates in Split-plot Designs	**5b. GRANT NUMBER**
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	**5d. PROJECT NUMBER**
Benjamin M. Lee, Capt, USAF	**5e. TASK NUMBER**
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S)	**8. PERFORMING ORGANIZATION REPORT NUMBER**
Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Street, Building 641 WPAFB OH 45433-7765	AFIT/OR-MS-ENS-10-06
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	**10. SPONSOR/MONITOR'S ACRONYM(S)**
INTENTIONALLY LEFT BLANK	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)**

12. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

13. SUPPLEMENTARY NOTES

14. ABSTRACT For any acquisition program, whether Department of Defense (DOD) or industry related, the primary driving factor behind the success of a program is whether or not the program remains within budget, stays on schedule and meets the defined performance requirements. If any of these three criteria are not met, the program manager may need to make challenging decisions. Typically, if the program is expected to not stay within budget or is expected to be delayed for one reason or another, the program manager will tend to limit areas of testing in order to meet these criteria. The result tends to be a reduction in the test budget and/or a shortening in the test timeline, both of which are already lean. The T&E community needs new test methodologies to test systems and gain insight on whether a system meets performance standards, within the budget and timeline constraints. In particular, both fundamental and advanced aspects of experimental design need to be adapted.

The use of experiential design within DOD has continued to grow because of the needed adaptation. Many different types of experiments have been used. An experimental design that is often needed is one that involves a restricted randomization design such as a split-plot design. Split-plot designs arise when specific factors are difficult (or impossible) to vary, a frequent occurrence within the T&E community. However, split-plot designs have limitations on the estimation of the whole plot (hard to change) and sub plot (easier to change) errors without the conduct of a sufficient number of replications for the design.

Within the timeline constraints for particular programs, sufficient replications are difficult, even impossible to complete. The inability to conduct the sufficient replications often lead to models that lack precision in error estimation and thus imprecision in corresponding conclusions.

This work develops and examines a methodology for analyzing test results conducted by split-plot designs using re-sampling techniques to provide better estimates of the error terms. The premise is to determine a set of rules using bootstrapping, a particular re-sampling technique, that can be applied to the analysis of a split-plot design, in order to create a representative regression model that can be used by the T&E community to gain required system insight.

15. SUBJECT TERMS
design of experiments, split-plot, bootstrap, re-sampling

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Dr Raymond Hill (ENS)
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636, ext 7469
U	**U**	**U**	**UU**	243	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18